National Park Service
U.S. Department of the Interior

Natural Resource Program Center

Data Management Plan for the North Coast and Cascades Network Inventory and Monitoring Program (2005)

Natural Resource Report NPS/NCCN/NRR—2009/078

ON THE COVER
Royal Basin, Olympic National Park.
Photograph by: Courtesy of NPS files

Data Management Plan for the North Coast and Cascades Network Inventory and Monitoring Program (2005)

Natural Resource Report NPS/NCCN/NRR—2009/078

John R. Boetsch
National Park Service
Olympic National Park
600 East Park Avenue
Port Angeles, WA 98362-6757

Bret Christoe
National Park Service
Mount Rainier National Park
55210 238th Ave. South
Ashford, WA 98304-9751

Ronald E. Holmes
National Park Service
North Cascades National Park Service Complex
810 State Route 20
Sedro Woolley, WA 98284-1239

February 2009

U.S. Department of the Interior
National Park Service
Natural Resource Program Center
Fort Collins, Colorado

The Natural Resource Publication series addresses natural resource topics that are of interest and applicability to a broad readership in the National Park Service and to others in the management of natural resources, including the scientific community, the public, and the NPS conservation and environmental constituencies. Manuscripts are peer-reviewed to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and is designed and published in a professional manner.

Natural Resource Reports are the designated medium for disseminating high priority, current natural resource management information with managerial application. The series targets a general, diverse audience, and may contain NPS policy considerations or address sensitive issues of management applicability. Examples of the diverse array of reports published in this series include vital signs monitoring plans; monitoring protocols; "how to" resource management papers; proceedings of resource management workshops or conferences; annual reports of resource programs or divisions of the Natural Resource Program Center; resource action plans; fact sheets; and regularly-published newsletters.

Printed copies of reports in these series may be produced in a limited quantity and they are only available as long as the supply lasts. This report is also available from the Natural Resource Publications Management website (http://www.nature.nps.gov/publications/NRPM) and the North Coast and Cascades Network Inventory and Monitoring website (http://science.nature.nps.gov/im/units/NCCN) on the Internet, or by sending a request to the address on the back cover.

Please cite this publication as:

Boetsch, J. R., B. Christoe, and R. E. Holmes. 2009. Data management plan for the North Coast and Cascades Network Inventory and Monitoring Program (2005). Natural Resource Report NPS/NCCN/NRR—2009/078. National Park Service, Fort Collins, Colorado.

The content of this document has not changed since the original version (2005); only styles and formatting have been updated to meet the Natural Resource Publication Series standards (2009).

NPS D-48, February 2009

Contents

Figures

Page

Executive Summary

The central mission of the National Park Service (NPS) Inventory and Monitoring Program is to provide timely and usable scientific information about the status and trends of park resources to park managers. To meet this challenge, we need an information management system that can effectively produce, maintain and distribute the products of scientific work done in our parks.

Good data management is the means by which a thorough understanding of the value of scientific information about our natural resources can become a part of our National Park Service heritage. Data management refers to the framework by which data are acquired, maintained, and made available. Data management is not an end unto itself, but instead is the means of maximizing the quality and utility of our natural resource information. A robust system for data management is particularly important for long-term programs where the lifespan of a data set will span the careers of several scientists. Seen in this way, it becomes obvious that data management is vital to the success of any long-term monitoring initiative.

The overarching goal of our data management system is to provide timely and usable scientific information about the status and trends of park resources to park managers. The success of our program hinges upon our ability to produce, manage and deliver this information to its intended audience. Our strategy for achieving this goal can be summarized as follows: Ensure the quality, interpretability, security, longevity and availability of our natural resource data. In implementing a data management system we will strive for the following:

- Confidence in the security and availability of natural resource data and related information
- Easy access to most information, and appropriate safeguards for sensitive information
- Awareness of the intended use and limitations of each data set
- Infrastructure and documentation that encourages data exploration
- Compatibility of data sets for exploration and analysis at larger scales and across disciplines
- Implementation of standards and procedures that facilitate information management, and that reinforce good habits among staff at all levels of project implementation – project leaders, technicians, and volunteer data collectors
- A proper balance between the standards needed to ensure quality and usability, and the flexibility to meet specific needs and encourage innovation
- A natural resource culture which views data not as a commodity but as the lifeblood of our work

The North Coast and Cascades Network (NCCN) Data Management Plan outlines how we intend to implement and maintain a system that will serve the data and information management needs of our Inventory and Monitoring Program. This plan reflects our commitment to establishing and maintaining a robust system for data management to ensure the availability and usability of high-quality natural resource information.

The NCCN Data Management Plan describes how our Network will:

- support Inventory and Monitoring Program objectives
- acquire and process data
- assure data quality
- document, analyze, summarize and disseminate data and information
- maintain nationally developed data management systems
- maintain, store and archive data

Acknowledgements

Much of the material in this plan was developed during a lively and productive period of collaboration among data managers from across the National Park Service. This level of collaboration makes it difficult to give proper credit to those who made significant contributions, and whose material appears in this document in one form or another. Nevertheless, we owe a debt of gratitude for the hard work and insight of those who participated in this process: Sara Stevens, Margaret Beer, Deborah Angell, Dorothy Mortenson, Doug Wilder, Gareth Rowell, Michael Williams, Geoffrey Sanders, Rob Daley, Teresa Leibfreid, Bill Moore, Patrick Flaherty, Fred Dieffenbach, and Velma Potash. We would also like to acknowledge Gary Entsminger for providing editorial revisions of some of the chapters drafted by this work group, and Joe Gregson and Steve Fancy for their comments and review of early drafts of the collaborative material. Additional credits are provided at the end of each chapter.

Finally, thanks to Roger Hoffman, Darin Swinney, Katherine Beirne, Anne Braaten, Lise Grace, Natalya Antonova and Ruth Jenkins for their many contributions to this plan.

Chapter 1 – Introduction

The central mission of the NPS Inventory and Monitoring Program is to provide timely and usable scientific information about the status and trends of park resources to park managers. To meet this challenge, we need an information management system that can effectively produce, maintain and distribute the products of scientific work done in our parks.

Good data management is the means by which a thorough understanding of the value of scientific information about our natural resources can become a part of our National Park Service heritage. Data management refers to the framework by which data are acquired, maintained, and made available. Data management is not an end unto itself, but instead is the means of maximizing the quality and utility of our natural resource information. A robust system for data management is particularly important for long-term programs where the lifespan of a data set will span the careers of several scientists. Seen in this way, it becomes obvious that data management is vital to the success of any long-term monitoring initiative.

This chapter summarizes the data management goals and objectives for our NCCN data management system. This Data Management Plan (DMP) documents the overarching strategy for ensuring that program data are documented, secure, and remain accessible and useful for future generations. The DMP, in turn, refers to other guidance documents and standard operating procedures which convey the specific standards and steps for achieving our data management goals.

1.1 – Inventory and Monitoring Program Overview

The Inventory and Monitoring (I&M) Program represents a long-term commitment by the National Park Service (NPS) to assess and document the status and trends of park ecological resources. In 1998, the National Parks Omnibus Management Act established the framework for the I&M Program to fully integrate natural resource monitoring and other scientific activities into the management processes of the National Park System.

The Act charges the Secretary of the Interior to "continually improve the ability of the National Park Service to provide state-of-the-art management, protection, and interpretation of and research on the resources of the National Park System", and to "… assure the full and proper utilization of the results of scientific studies for park management decisions." Section 5934 of the Act requires the Secretary of the Interior to develop a program of "inventory and monitoring of National Park System resources to establish baseline information and to provide information on the long-term trends in the condition of National Park System resources."

National Park Service Mission
The National Park Service preserves unimpaired the natural and cultural resources and values of the national park system for the enjoyment, education and inspiration of this and future generations...

To carry out this mission, the National Park Service initiated a service-wide, natural resource Inventory and Monitoring Program encompassing 270 parks with significant natural resources. Ecologically similar parks among the 270 chosen were grouped into 32 networks. Each I&M Network has been tasked with documenting existing park

vertebrates and vascular plants, developing a management based, ecological monitoring program with a written plan and protocols, and a Data Management Plan that describes our approach to network information management.

1.2 – Data Management Goals and Objectives

The overarching goal of our data management system is to provide timely and usable scientific information about the status and trends of park resources to park managers. The success of our program hinges upon our ability to produce, manage and deliver this information to its intended audience. Specifically, our goal is to develop and maintain a system for producing value-added natural resource data. This system should ensure that natural resource information are available, usable, sharable, integrated and interpreted:

- *Available* – We have the means of knowing that the data exist and where the data can be found.
- *Usable* – Data are stored in a stable, reliable, and interpretable data retrieval system.
- *Shareable* – Data products are complete, certified for quality assurance, screened for sensitive information, formatted for use and documented for interpretation by others.
- *Integrated* – Data products are consistent with data exchange standards, interoperable with related natural resource data sets, and are collected and stored in a way that optimizes the tradeoff between meeting local needs and achieving compatibility with other agencies and partners.
- *Interpreted* – The data have been reviewed, summarized and transformed into useful information.

Although our primary responsibility is for ecological data and related information derived from Inventory and Monitoring Program activities, it is anticipated that the scope of our data management efforts will expand to include all significant natural resource data specific to the North Coast and Cascades Network and the parks of which it is composed.

In implementing a data management system, we will strive for the following:

- Confidence in the security and availability of natural resource data and related information
- Easy access to most information, and appropriate safeguards for sensitive information
- Awareness of the intended use and limitations of each data set
- Infrastructure and documentation that encourages data exploration
- Compatibility of data sets for exploration and analysis at larger scales and across disciplines
- Implementation of standards and procedures that facilitate information management, and that reinforce good habits among staff at all levels of project implementation – project leaders, technicians, and volunteer data collectors
- A proper balance between the standards needed to ensure quality and usability, and the flexibility to meet specific needs and encourage innovation
- A natural resource culture which views data not as a commodity but as the lifeblood of our work

1.3 – Data and Data Management – Providing Context

Collection of natural resource data is our first step toward understanding the ecosystems within our national parks. These ecosystems are evolving, as is our knowledge of them and how they function. We use these "raw" data to analyze, synthesize, and model aspects of ecosystems. In turn, we use our results and interpretations to make decisions about the Park's natural resources. Thus, *data* collected and maintained by the North Coast and Cascades Network will become *information* through analysis, synthesis, and modeling. Information is the common currency among those involved in stewardship projects throughout our National Park System.

Any good set of data – whether collected last week or 20 years ago – must provide enough information about itself that it can be preserved and used appropriately. Anyone using these data will need to know as much as possible about how and why these data were collected. Therefore, our data management system cannot simply attend to the tables, fields, and values that make up a data set. It must also provide a process for developing, preserving, and integrating the context that makes data valuable and interpretable. Although this means more time will be spent documenting data sets, it will result in better preservation and presentation of data.

We sometimes use the term "data" in a way that also encompasses other products that are generated alongside primary tabular and spatial data. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 1.1). To meet I&M Program goals, and to ensure adequate context for the primary data products, these categories of project deliverables all require some level of management to ensure their quality and availability. We intend to integrate the manner in which our network creates, manages, and makes available the products of our scientific efforts. Thus, we will take a holistic view of how natural resource data are generated, processed, finalized, and provided to others. All phases of data and information processing are integrated, and information about each phase is shared through good documentation.

Table 1.1. Categories of data products and project deliverables

Category	Examples
Raw data	GPS rover files, field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens
Compiled/derived data	Relational databases, tabular data files, GIS layers, maps, species checklists
Documentation	Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionary, FGDC/NBII metadata, data design documentation, quality assurance report, catalog of specimens/photographs
Reports	Annual progress report, final report (technical or general audience), periodic trend analysis report, publication
Administrative records	Contracts and agreements, study plans, research permits/applications, other critical administrative correspondence

1.4 – Sources of Natural Resource Data

There are many potential sources of important information about the condition of natural resources in our parks. The types of work that may generate natural resource data about park resources include:

- Inventories
- Monitoring
- Protocol development pilot studies
- Special focus studies done by internal staff, contractors or cooperators
- External research projects
- Monitoring or research studies done by other agencies on park or adjacent lands
- Resource impact evaluations related to park planning and compliance with regulations
- Resource management and restoration work

Prioritizing data management efforts in a sea of unmanaged data

Our highest priority is to produce and curate high-quality, well-documented data originating from the Inventory and Monitoring Program. As time and resources permit, we will work toward raising the level of data management for current projects, legacy data, and data originating outside the Inventory and Monitoring Program.

1. We will place greatest emphasis on projects that are just beginning to be developed and implemented, because inserting good data management practices from the very beginning is most effective.
2. Updating data management for ongoing projects and conversion of legacy data will be prioritized according to the value of the information they provide.

Because the I&M Program focuses on long-term monitoring and natural resource inventories, our first priority should be toward the data and information that we derive from these primary efforts. However, we can easily apply the same standards, procedures, infrastructure and attitudes about data management to other natural resource data sources. One challenge will be to prioritize and manage workload and other resources. Naturally, high-profile data sets that provide crucial information to park management will be prioritized for data management regardless of funding source.

This plan covers four major categories of data that are coordinated or managed by NCCN. These are:

1. *Data managed in service-wide databases.* NCCN uses three data systems developed by the I&M WASO office. *NatureBib* is used as a bibliographic tool for cataloging reports, publications, or other documents that relate to natural resources in park units. *Dataset Catalog* is used to document primarily non-spatial databases or other data assemblages. *NPSpecies* is used to develop and maintain lists of vertebrates and vascular plants in network parks, along with associated supporting evidence.

2. *Data developed or acquired directly by the network as a result of inventory, monitoring, or other projects*. This category includes project-related protocols, reports, spatial data, and associated materials such as field notes and photographs provided to NCCN by contractors or developed by NCCN staff. Projects can be short-term (i.e., one to two years duration) or long-term (e.g., ongoing monitoring).

3. *Data that, while not developed or maintained by NCCN, are used as data sources or provide context to other data sets.* Examples of this category include: base GIS data

developed by parks, other agencies or organizations; national or international taxonomic classification systems; climate or hydrologic data collected by regional or national entities; data and products developed under the auspices of a research permit.

4. *Data acquired and maintained by network parks that NCCN assists in managing.* Because of the lack of data management expertise in many network parks, NCCN provides data management assistance for high-priority data sets or those that may benefit from standardized procedures. Examples include: park observations databases for wildlife and rare plant data; legacy natural resource monitoring data sets; and data on exotic invasive plant species.

These categories can contain one or more of the following data formats:

- hard-copy documents (e.g., reports, field notes, survey forms, maps, references, administrative documents)
- objects (e.g., specimens, samples, photographs, slides)
- tabular data (e.g., databases, spreadsheets, tables, delimited files)
- spatial data (e.g., shapefiles, coverages, remote-sensing data)
- other digital files (e.g., Word files, email, websites, digital images)

Each of these data formats has specific requirements for ongoing management and maintenance, which we address in this plan.

1.5 – Makeup of the North Coast and Cascades Network

The North Coast and Cascades Network (NCCN or Network) is composed of eight park units in western Washington and the northwestern corner of Oregon. They include five small, historically based parks and three larger, mountainous parks that are coastal, continental, or both. Seven of these parks contain significant natural resources, and thus are the subject of this plan (Figure 1.1):

- Ebey's Landing National Historical Reserve (EBLA)
- Fort Vancouver National Historic Site (FOVA)
- Lewis and Clark National Historical Park (LEWI), formerly Fort Clatsop National Memorial
- Mount Rainier National Park (MORA)
- North Cascades National Park Service Complex (NOCA)
- Olympic National Park (OLYM)
- San Juan Island National Historical Park (SAJH)

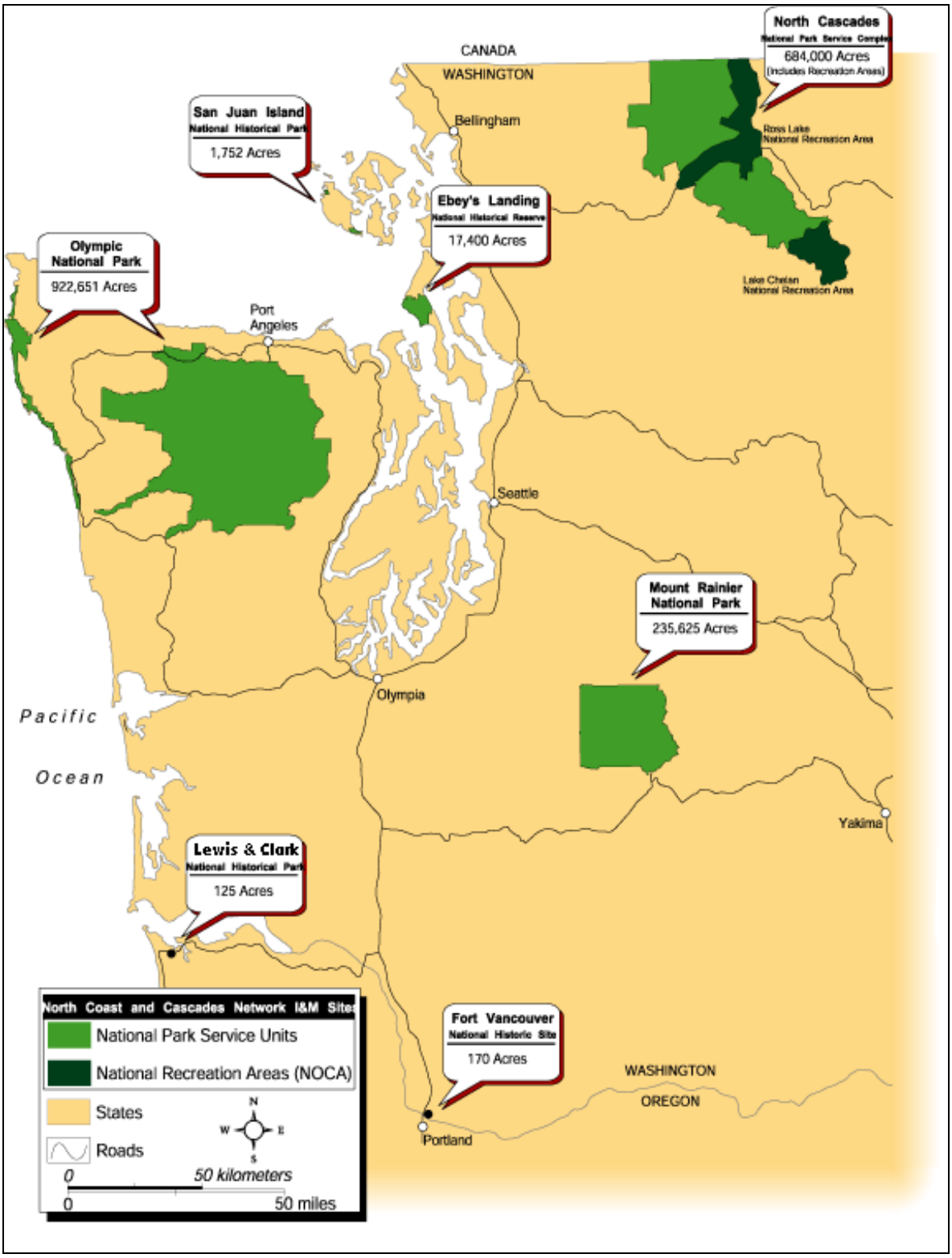

Figure 1.1. Parks with significant natural resources in the North Coast and Cascades Network

NCCN is unique in being the only NPS network that contains two parks with prototype I&M programs – one at Olympic National Park and another at North Cascades National Park Service Complex. Although these parks were competitively selected as prototypes in 1993, they did not receive prototype funding until 2000, simultaneous with network funding. Subsequent to funding as separate entities, the management of these prototype programs was united in that funding and implementation decisions are made collectively. Funding from the prototype programs supports scientific and data management personnel stationed at each prototype, which augments the Network Coordinator and Data Manager positions funded by the network. The mission of the prototype programs is to function in support of the network and its constituent parks by providing guidance and expertise related to monitoring design, data management, analysis, and reporting – while still meeting their individual prototype responsibilities as national centers of innovation and excellence.

Board of Directors

The Board of Directors is composed of park superintendents of the seven park units that comprise NCCN. The Pacific West Region I&M Coordinator and the Network Coordinator are *ex-officio* members of the Board. The primary role of the Board is to promote accountability and effectiveness for the I&M Program by reviewing progress and quality control for the network, and to oversee spending of network funds. The Board also provides guidance and oversight for the Technical Committee in the design and implementation of vital signs monitoring and other activities related to the Natural Resource Challenge. An important role of the Board is to seek strategies for leveraging network funds and personnel to best accomplish the inventory and monitoring and other natural resource needs of network parks, and for finding additional funding sources. As the governing body of the network, the Board is consulted on hiring of new personnel using funds provided to the network in combination with base funds and other sources. There will be at least two formal board meetings annually, though any member can call a meeting as needed.

Technical Committee

The network Technical Committee is composed of the chiefs of natural resources for the member parks, key natural resource staff and subject area specialists, a science advisor, a liaison from USGS-Biological Resources Division, and the GIS and data management staff for the member parks. This committee meets monthly to coordinate inventory and monitoring activities and to propose decision points to the Board of Directors. The Network Coordinator convenes and chairs meetings of the Technical Committee. The Technical Committee also provides general oversight and guidance for work groups that are responsible for topic-specific activities.

Discipline work groups

Discipline work groups are made up of scientists and resource managers from network parks who are working together to manage and coordinate inventory and monitoring within the network. These work groups are defined by the broad topical categories for which they are responsible: Aquatic (water quality, aquatic invertebrates, amphibians, fish), Marine/Coastal (estuarine/intertidal communities), Geology (bedrock, landforms, glaciers, erosion), Terrestrial Wildlife (mammals, birds, reptiles, terrestrial invertebrates), Vegetation (vascular, non-vascular

plants, exotic plant management, rare plants), and Atmospherics (climate, atmospheric deposition). Each work group meets as needed and meetings are coordinated by a chairperson appointed by each group. Meeting agendas and minutes are posted on the network's intranet along with any other resulting documentation such as budgets, assignments, and reports.

Data management work group

This work group consists of the three data managers of the network (two prototype data managers and one network data manager), and the GIS staff of the network parks and regional support office. This group meets to develop and evaluate strategies for maintaining our data management system. Because good data management is a collaborative endeavor, the work group will work toward increasing the awareness and competency for good data management among park scientists, resource managers, field technicians, and others who produce, maintain or use natural resource data. The group meets at least quarterly to communicate new methods, procedures and standards, and to set priorities for data management within the network. The group also designates a rotating chairperson for 4-month terms to represent the data management group at technical committee meetings and discipline work group meetings.

1.6 – Scope and Organization of the Data Management Plan

The North Coast and Cascades Network Data Management Plan outlines how we intend to implement and maintain a system that will serve the data and information management needs of our Inventory and Monitoring Program. This plan reflects our commitment to establishing and maintaining a robust system for data management to ensure the availability and usability of high-quality natural resource information.

The NCCN Data Management Plan describes how our Network will:

- support I&M Program objectives
- acquire and process data
- assure data quality
- document, analyze, summarize and disseminate data and information
- maintain nationally developed data management systems
- maintain, store and archive data

We have developed this data management plan in modular format in the hope of increasing readability and usability among a wider audience. This format allows users to more easily locate and access subdocuments pertaining to a particular element of information management guidance or process. Individual procedural documents within the plan can be provided as standalone documents to park and regional data management staff, cooperators and project leaders. This format also allows for a more efficient review as technical experts can be provided with applicable portions of the plan based on their expertise (GIS, databases, metadata, GPS, etc.). There are three document types that make up this plan:

- NCCN Data Management Plan
- Guidance Documents

- Standard Operating Procedures

Note: These documents describe the data management system and practices as a whole, whereas project-specific details are described in the standard operating procedures included in project protocols.

This Data Management Plan is composed of chapters that provide overview information that is more stable, whereas information that is more subject to change is more typically found in Guidance Documents and Standard Operating Procedures (SOPs). The conceptual model in Figure 1.2 shows the hierarchy of documentation related to NCCN information management.

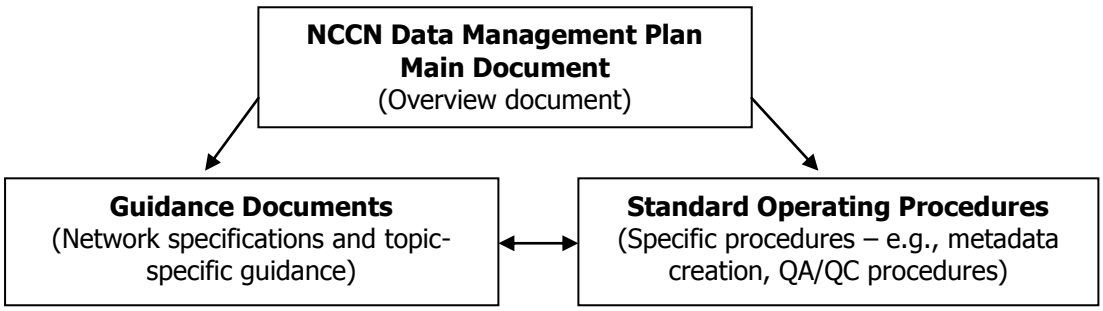

Figure 1.2. Hierarchy of NCCN data management documents

Standard Operating Procedures (SOP) describe work processes that are to be followed by network staff and cooperators. SOPs are useful in that they document the way activities are to be performed to facilitate consistent conformance to technical and quality assurance requirements. The development and use of SOPs promotes quality through consistent implementation of a process or procedure within the organization, even if there are temporary or permanent personnel changes.

1.7 – Guidance Documentation Associated with this Plan

Once completed, the following documents will be made available on the NCCN website under Data Management (http://science.nature.nps.gov/im/units/nccn/datamgmt_guide.cfm). Specific procedures and SOPs will typically be included in these documents as appendices.

- Tracking Project Information
- Naming Conventions Files and Data Objects
- Data Design Guidelines and Database Specifications
- Reporting Guidelines and Templates
- Specifications for Data Products and Project Deliverables
- GPS Data Collection and Processing
- GIS Development and Management Guidelines
- Template Language for Contracts and Agreements
- Guidelines for Managing Third-party Data
- Standards for Voucher Collection and Cataloguing
- Photographic Image Management Strategy

- Quality Assurance Guidelines
- Metadata Procedures and Specifications
- NPSpecies Database Management Procedures
- NatureBib Database Management Procedures
- Guidelines for Data Distribution and Data Discovery
- Sensitive Information Management Guidelines
- Archival Procedures for Project Information
- Webpage Development and Management Guidelines

1.8 – Review and Revisions

The NCCN Data Management Plan covers I&M Program needs based on the most current information systems technology relevant through 2005. This document will undergo an initial peer review by members of the NPS data management community, the NCCN Technical Committee, and informal review by individuals from other agencies. Following peer review, maintenance of this plan will be incremental as changes are needed to reflect best practices and current information.

Revisions to this plan and associated data management documents will be made as needed, and the whole plan will be reevaluated every 3 to 5 years. The data management work group for the network will be responsible for ongoing maintenance of data management documents. Recommendations for changes can be forwarded to one of the data managers by any interested party or consumer of network inventory and monitoring data – park resource managers, project managers, technicians, superintendents, external users, etc. These recommendations will be brought up and discussed at the regular meetings of the data management work group. Revisions will be tracked using the revision log at the beginning of the document. Simple changes can be made immediately in the document, and more complex changes and decision points can be recorded for subsequent plan updates.

Credits

The content of this chapter was adapted from material developed by Sara Stevens (Northeast Coast and Barrier Network) and Doug Wilder (Central Alaska Network).

Chapter 2 – Data Stewardship Roles and Responsibilities

Data management is a complex process characterized as much by attitudes and habits as it is by infrastructure, standards and procedures. Although primary responsibility resides with the data managers, good data stewardship could not possibly be accomplished by data managers alone – it is truly a collaborative endeavor that involves many people with a broad range of tasks and responsibilities. As such, a valid data management system must be developed and continually modified to meet the needs of everyone with a role in coordinating, generating, maintaining, and using natural resource information in its many forms. This is a diverse group made up of park managers, scientists, technical specialists, and interpretive staff. A successful data management system is maintained by reinforcing communication and awareness among those responsible for the origin, quality, disposition, and use of the data.

The demand for detailed, quality information about natural resources requires a group of people working together to steward data and information assets. The complex and technical character of ecological data, combined with the diverse and constantly evolving information technology, requires knowledgeable individuals from many areas to come together to ensure that data are collected using appropriate methods, and that resulting data products are well managed. Natural resource information must be credible, representative, and available for current and future needs.

2.1 – Roles and Responsibilities

A *role* is a function or position (e.g., Project Leader)
A *responsibility* is a duty or obligation (e.g., review data records)

Table 2.1 summarizes the roles and responsibilities of various personnel. These roles are listed 'from the ground up' to help demonstrate the hierarchy of responsibilities. For example, a project leader is ultimately responsible for the activities listed in the field level roles of crew leader and crew member. Also, the network coordinator ensures that the network data manager and ecologist achieve the required performance level.

Table 2.1. Roles and responsibilities for data stewardship

Role	Data Stewardship Responsibilities
Project Crew Member	Collect, record, and verify data
Project Crew Leader	Supervise crew and organize data
GIS Specialist or Data Technician	Process and manage data
Information Technology Specialist	Provide IT support for hardware, software, networking
Project Leader	Oversee project operations, including data management
Resource Specialist/Ecologist	Validate and make decisions about data. Integrate science in park and network activities.
GIS Coordinator	Support park management objectives with GIS and resource information management
Data Manager	Develop and support data management system. Ensure project data are organized, compliant, safe, and available.
Database Application Developer	Know and use database software and database applications
Curator	Oversee all aspects of specimen acquisition, documentation, and preservation, and manage the park collections
Statistician or Biometrician	Analyze data and/or consult on analysis
Network/Prototype Coordinator	Coordinate and oversee all network activities
Park Research Coordinator	Facilitate data acquisition by external researchers. Communicate NPS requirements to permit holders.
I&M Data Manager (National Level)	Provide Service-wide database availability and support
End Users (managers, scientists, interpreters, public)	Inform the scope and direction of science information needs and activities. Interpret information and apply to decisions.

Chief personnel involved with data management include Project Leaders and Data Managers. The Network Coordinator interacts with Project Leaders to ensure that timelines for data entry, validation, verification, summarization/analysis and reporting are met. Figure 2.1 illustrates the core data management duties of the Project Leader and Data Manager and where those duties overlap.

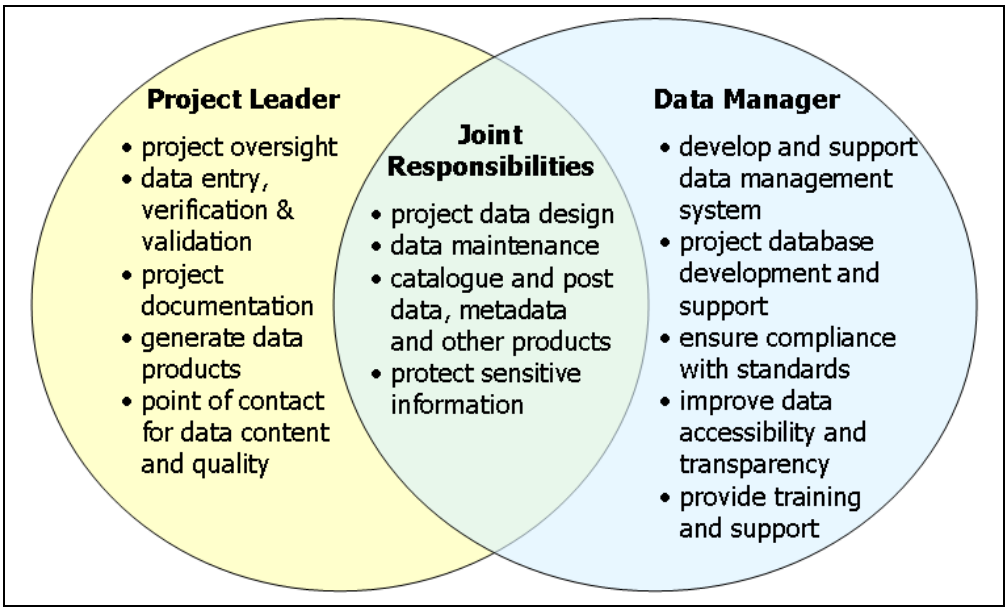

Figure 2.1. Core project data stewardship duties of project leaders and data managers

2.2 – Data Management Coordination

The Natural Resource Challenge states that collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in national parks. The I&M Program encourages coordination among participants at all levels to help ensure that data collected in parks are properly developed, maintained, and made available for management decision-making, research, and education.

The network data managers work with I&M Program data management staff and regional resource information management personnel to maintain a high-level of involvement in service-wide and regional databases and data management policy. The data managers also work locally with network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in the compatibility and availability of data sets.

Key contacts include park GIS specialists and the project leaders for each monitoring or inventory project. Involvement and input from park scientists and resource management staff is also essential. We rely on everyone in the network for the successful development of planning materials, inventory study plans, and monitoring protocols. Consistent and productive communication among these individuals leads to common understanding and better synchronization of data management activities. This communication may take the form of personal visits, phone calls, email, joint meetings and training sessions, and participation in meetings and work of NCCN's Technical Committee and Board of Directors.

2.3 – Cooperation with Other Networks

Data managers throughout NPS regularly coordinate with each other and national program staff via annual meetings, conference calls, workgroups, a listserv, web sites, and informal communication. Data managers have already demonstrated effective cooperation in the ways that we share ideas and technology, and collaborate to develop data management standards and documentation. NCCN data managers will maintain an active role in promoting practical levels of compatibility among protocols and data sets involving other networks and organizations.

2.4 – Project Stewardship

Since the implementation of each project will normally require the expertise and involvement of several people over a period of months or years, it makes sense that one person is charged with keeping track of the objectives, requirements, and progress for each project. This project leader is usually a park scientist or resource management specialist with training and experience in the field of science related to the inventory or monitoring project. This expertise helps the project leader to oversee data acquisition and processing, coordinate with GIS and data managers on information management needs, and communicate project objectives and results. A project leader who cannot act as an authority for the data should work with the appropriate specialists to account for those aspects of data stewardship.

To ensure the quality of each project, including data requirements, project leaders should be assigned only those projects they can effectively oversee based on workload and other relevant factors. Unless the project is of a very short duration (3 months or less), each project should have at least one alternate project leader to provide continuity in case the principal project leader becomes unavailable.

2.4.1 – Data Stewardship: Sharing Responsibilities

Keeping track of data from the time of acquisition until it is no longer useful is the shared responsibility of everyone involved with data – producers, analysts, managers, and end users. This, in essence, is *data stewardship*. It is a principle of mutual accountability rather than one particular job for one individual. Successful data stewardship requires that people involved in NCCN activities learn and understand the expectations for continuous data management. This is equally important for NCCN staff, park employees, and contractors or cooperators. All project participants should receive training, briefings, materials, and additional regular communication about data stewardship from project leaders, and data managers. The purpose is to promote the appropriate level of understanding about how their efforts relate to park and network management objectives, NPS policies, and other federal government requirements.

2.4.2 – Documentation is Key

If one shared responsibility stands above the pack in importance and value, it is the careful documentation of data sets, the data source(s), and the methodology by which the data were collected or acquired. This documentation establishes the basis for the appropriate use of the data in resulting analysis and products, both in the short-term and long-term. Network monitoring

protocols contain key elements of data documentation, although additional documentation will need to be provided periodically. See the Chapter 7 of this plan for important guidance and reference for documentation and metadata.

2.4.3 – The Hub of Data Stewardship

Project leaders, data managers, and GIS specialists comprise the central data management team for inventory and monitoring projects. Each is responsible for certain aspects of project data, and all share responsibility for some overlapping tasks. Because of the collaborative nature of project data management, good communication among project leaders, data managers, and GIS specialists is essential to meeting program goals.

2.4.4 – Project Stewards

The project leader is responsible for data quality during all phases of the project, including collecting, entering, handling, reviewing, summarizing, and reporting data. Developing project documentation and metadata are crucial elements of the project leader's role. Figure 2.1 illustrates some of the overlap between project and data management responsibilities.

Specifically, a project leader is responsible for:

- ***project documentation*** that describes all aspects of project implementation
- development and implementation of ***standard procedures*** for field data collection and handling
- ***quality assurance and quality control measures***, which include the supervision and certification of all field operations, staff training, equipment calibration, species identification, data collection, data entry, verification, and validation
- maintaining concise explanatory documentation of all deviations from standard procedures
- providing detailed documentation for each field data collection period
- maintaining hard copies of data forms and archiving of original data forms
- scheduling of regular project milestones such as data collection periods, data processing target dates, and reporting deadlines
- regular summary reports and periodic trend analysis
- acting as the main point of contact concerning data content

The project leader will also work closely with the data manager to:

- develop quality assurance and quality control procedures specific to project operations
- identify training needs for staff related to data management philosophy, database software use, and quality control procedures
- coordinate changes to the field data forms and the user interface for the project database
- document and maintain master data
- identify sensitive information that requires special consideration prior to distribution
- manage the archival process to ensure regular archiving of project documentation, original field data, databases, reports and summaries, and other products from the project

- define the process for transforming raw data into meaningful information
- create data summary procedures to automate and standardize this transformation process
- identify, prioritize, and convert legacy data sets to modern formats
- increase the interpretability and accessibility of existing natural resource information

2.4.5 – Data Managers

The data manager is responsible for ensuring the compatibility of project data with program standards, for designing the infrastructure for project data, and for ensuring long-term data integrity and availability.

Data managers will:

- ***develop and maintain the data management system*** for metadata creation and project data management
- ***create and support project databases*** in accordance with best practices and current standards
- ***provide training*** in the theory and practice of data management as tailored to the needs of project personnel
- develop ways to ***improve the accessibility and transparency*** of digital data
- establish and implement procedures to protect sensitive data according to project needs
- collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives

Data managers will also work closely with the project leader to:

- define the scope of the project data and create a data design that meets project needs
- become familiar with how project data are collected, handled, and used
- review quality assurance aspects of project protocols and standard procedure documents
- identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick-lists and conditional validation rules
- create a user interface that streamlines the process of data entry, review, validation, and summarization that is consistent with the capabilities of the project staff
- develop automated database procedures to improve the data summarization and reporting process
- make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data
- ensure regular archiving of project materials
- inform project staff of changes and advances in data management practices

2.4.6 – GIS Specialists

Park GIS specialists play a crucial supporting role for NCCN projects. GIS support includes project planning to determine the GIS data and analysis needs for a project. GIS specialists will provide consultation to project leaders on field collection of spatial data including the use of GPS

and other spatial data collection techniques. They will also coordinate importing of spatial data into the GIS, work with project leaders to analyze spatial data, and provide the results in map or tabular form. The GIS specialists will work with project leaders to properly document data in compliance with spatial metadata standards and will be responsible for stewardship of GIS data and products.

GIS specialists will also work directly with data managers to design databases and applications. GIS specialists will work closely with the data managers to create relationships between GIS and non-spatial data and create appropriate database and GIS applications to facilitate the analysis of both spatial and non-spatial data.

The Columbia Cascades Support Office (CCSO) in Seattle will continue their role as primary GIS support for EBLA, FOVA, LEWI, and SAJH since these parks do not have GIS Specialists on staff. However, GIS specialists from MORA, NOCA, and OLYM may also provide GIS support to NCCN projects taking place in these parks.

Credits

This chapter was adapted from concepts and material developed in collaboration with Rob Daley (Greater Yellowstone Network).

Chapter 3 – Data Management Resources: Infrastructure and Systems Architecture

Our computer resource infrastructure is composed of computers and servers that are functionally or directly linked through computer networking services. This infrastructure represents the foundation upon which our network information system is built. Systems architecture signifies the applications, database systems, repositories, and software tools that make up the framework of our data management enterprise.

Our I&M program relies heavily on park, regional, and national IT personnel and resources to maintain the computer resource infrastructure. This includes but is not limited to hardware replacement, software installation and support, security updates, virus-protection, telecommunications networking, and backups of servers. Therefore communication with park and regional IT specialists is essential to ensure adequate resources and service continuity for our systems architecture. Rather than focusing on a detailed description of our current computer resources, this chapter will instead describe our infrastructure in more general terms and focus more specifically on the systems architecture that is central to data management.

3.1 – Computer Resources Infrastructure

An important element of a data management program is a reliable, secure network of computers and servers. Our digital infrastructure has three main components: park-based local area networks (LAN), network data servers, and servers maintained at the national level. This infrastructure is maintained by park, regional, and national IT specialists, who administer all aspects of system security and backups.

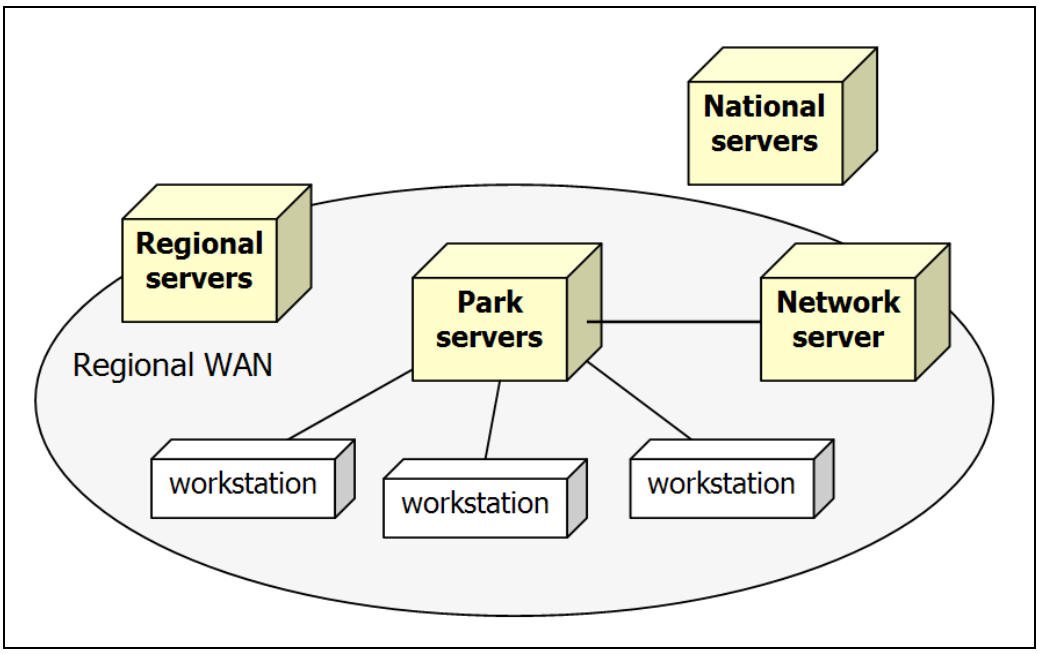

Figure 3.1. Schematic representing the logical layout and connectivity of computer resources

These components each host different parts of our natural resource information system.

National servers

- Master applications – integrated client-server versions of NatureBib, NPSpecies, NR-GIS Metadata Database
- Centralized repositories – NR-GIS Data Store, Protocol Clearinghouse
- Public access sites – portals to NatureBib, NPSpecies, NPSFocus, and websites for monitoring networks

Network data servers

- Master project databases – compiled data sets for monitoring projects and other multi-year efforts that have been certified for data quality
- Common lookup tables – e.g., parks, projects, personnel, species
- Project tracking application – used to track project status, contact information, product due dates
- Network digital library – network repository for finished versions of project deliverables for network projects (e.g., reports, methods documentation, data files, metadata, etc.)

Park LAN

- Local applications – desktop versions of national applications such as NPSpecies and Dataset Catalog
- Working files – working databases, draft geospatial themes, drafts of reports, administrative records
- Park digital library – base spatial data, imagery, and finished versions of park project deliverables
- Park GIS files – base spatial data, imagery, and project-specific themes

3.1.1 – Maintaining Digital Files

Each of the NCCN park LANs accommodates hierarchical directory structures for storing digital files. There are five main categories of directory structure sections in which digital files can be maintained (also see Chapter 10):

Admin	Documents related to program administration.
Databases	Local copies of national databases (e.g., NPSpecies), databases for common lookup tables, and project back-end databases are maintained. This section may also have separate subfolders for database development.
Libraries	Read-only storage of finished project deliverables, catalogued photographs, and other reference documents generated and maintained by the program.
Working	Workspace where groups and individuals can maintain draft material and other files as arranged by group then by fiscal or calendar year. The layout of folders and subfolders is more flexible here than elsewhere. On an annual basis, project leaders and administrators should work to clean out these sections by identifying material that belongs in one of the Libraries or elsewhere, deleting unneeded files, and moving the remainder to offline media or to a read-only folder named by year.

GIS — Base spatial data, imagery, and project-specific themes. The typical arrangement of folders and subfolders is park-specific. Some project-specific themes in development reside in working sections until they are migrated here upon completion. This section will typically be found on a separate drive due to different storage space requirements and backup issues, and because its scope is not confined to the I&M program.

The key aspects of this file management strategy are as follows:

1. Working files are kept separate from finished products.
2. Finished products are typically read-only, except for "inbox" folders where users can drop things off to be cataloged and filed.
3. Standards such as naming conventions and hierarchical filing are enforced within the Libraries, Database, and GIS sections. Although less stringent in other sections, these conventions are encouraged as good practice.

3.2 – National Information Management Systems

The need for effective natural resource information management cuts across NPS divisional boundaries. For this reason, information management strategies must be defined at the highest level possible. In this context, integrated inventory and monitoring of natural resources is multidisciplinary and requires national-level, programmatic data and information management strategies for success.

The basic strategy of natural resource and therefore inventory and monitoring information management is to provide integrated natural resource databases and information systems that enhance NPS managers' and staff's access and use of timely and valid data and information for management decisions, resource protection, and interpretation. Inventory and monitoring information needs are broadly separated into two categories:

- *Detailed data and information needed for onsite resource management and protection*. The information used to guide natural resource management decisions must be specific to inform and be useful to management staff at parks and central offices.
- *Summary information needed to describe the resources and their condition*. This kind of information usually needs to be aggregated across the National Park Service for use by NPS and DOI managers and central office personnel to answer requests from Congress and for budget, program, and project planning.

The NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and implement a national-level, program-wide information management framework. NRPC and I&M staff integrate desktop database applications with internet-based databases to serve both local and national-level data and information requirements. NRPC staff members work with regional and support office staff to develop extensible desktop GIS systems that integrate closely with the database systems. Centralized data archiving and distribution capabilities at the NRPC provide for long term data security and storage. NRPC sponsors training courses on data management, I&M techniques, and remote sensing to assist I&M data managers with developing and effectively utilizing natural resource information.

National-level application architecture

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NR-GIS Metadata Database) utilize a distributed application architecture with both desktop and internet-accessible (master) components (Figure 3.2).

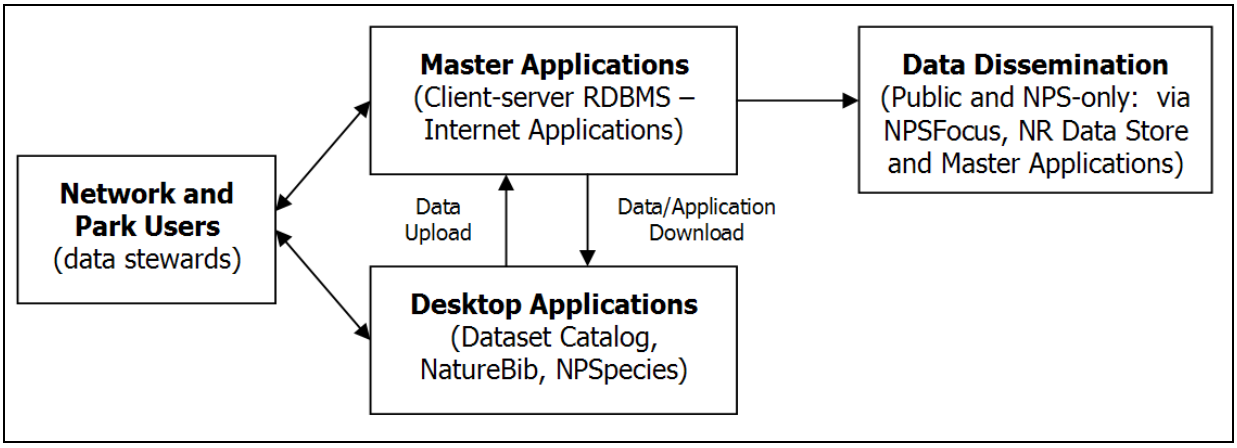

Figure 3.2. Model of the national-level application architecture

NatureBib
NatureBib is the master database for bibliographic references that merges a number of previously separate databases such as Whitetail Deer Management Bibliography (DeerBib), Geologic Resource Bibliography (GRBib), and others. It also contains citation data from independent databases like NPSpecies and the Dataset Catalog and NR-GIS Metadata Database. It currently focuses on natural resource references but may eventually be linked to references on cultural resources and other park operations. As with NPSpecies and NR-GIS Metadata Database, it is possible to download data from the master web version into the MS Access desktop version that can be used locally on computers with limited internet connectivity (*http://science.nature.nps.gov/im/apps/nrbib/*).

NPSpecies
NPSpecies is the master species database for the NPS. The database lists the species that occur in or near each park and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). Taxonomy and nomenclature are based on ITIS, the interagency Integrated Taxonomic Information System. The master version of NPSpecies for each park or network can be downloaded from the master website into an MS Access version of NPSpecies. The internet-based version is the master database, which can be accessed via password-protected logins administered by park, network and regional data stewards (points of contact or POCs) assigned for each park and network. The master database requires that species lists are certified by parks before any data will be available to the public. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park and will be linked to NR-GIS Metadata Database to document biological inventory

products. The MS Access application and additional details can be found at the NPSpecies website *(http://science.nature.nps.gov/im/apps/npspp/index.cfm).*

Dataset Catalog and NR-GIS Metadata Database
Dataset Catalog is a desktop metadata database application developed by the I&M Program to provide a tool that parks, networks, and cooperators can use to inventory and manage data set holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for cataloging abbreviated metadata about a variety of digital and non-digital natural resource data sets. The Dataset Catalog helps parks and networks begin to meet Executive Order 12906 mandating federal agencies to document all data collected after January 1995. It provides brief metadata and a comprehensive list of all resource data sets for use in data management, project planning, and more stringent metadata activities. As with other service-wide applications, the master metadata database (NR-GIS Metadata Database) is available through a website and will be linked to NPSpecies (the NPS species database) and NatureBib (the bibliographic database). A Microsoft Access version may be downloaded from the website (*Dataset Catalog: http://science.nature.nps.gov/im/apps/datacat/index.htm and NR-GIS Metadata Database: http://science.nature.nps.gov/nrdata).*

Other National-Level Inventory and Monitoring Information Management and GIS Applications

NPSTORET
STORET is an interagency water quality database developed and supported by the Environmental Protection Agency (EPA) to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. STORET is used by NPS as a repository of physical, chemical, biological, and other monitoring data collected in and around national park units by park staff, contractors, and cooperators. The NPS operates its own service-wide copy of STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be accessible to the public. NPS Director's Order 77 indicates that the NPS should archive water quality data in STORET, and the NPS Water Resources Division (WRD) requires that any data collected as part of a funded WRD project be archived in STORET. NPSTORET (also known as Water Quality Database Templates) the NPS master database designed to facilitate park-level standardized reporting for STORET. The database is still in development, but metadata, protocols, data dictionaries, and reporting capabilities are available through a front-end form. Upon implementation, network staff and cooperators will be able to use the MS Access version of NPSTORET either as a direct database for data entry and management, or as a means of submitting data for upload to STORET by WRD staff. The MS Access application and additional details can be found at: *http://www.nature.nps.gov/water/infodata.htm.* Additional information on STORET can be found at: *http://www.epa.gov/storet.*

Natural Resource Database Template
The Natural Resource Database Template (NRDT) is a flexible, relational database in MS Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database can be used as a standalone database or in conjunction with the GIS software (e.g., ArcView or ArcGIS) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended by different parks and networks depending on the components of their inventory and

monitoring program and the specific sampling protocols they use. Natural Resource Database Template is a key component of the I&M program's standardized monitoring protocols. These monitoring protocols include separate modules detailing different aspects of monitoring project implementation, from sampling design to data analysis and reporting, and include data management components that describe database table structure, data entry forms and quality checking routines. Approved monitoring protocols, including the databases that are based on the Database Template, are made available through a web-based protocol clearinghouse (see below). A description of the Database Template application, a data dictionary, and example implementations are located on the NR Database Template website (*http://science.nature.nps.gov/im/apps/template/index.cfm*).

Natural Resource Monitoring Protocols Clearinghouse
The Natural Resource Monitoring Protocol Clearinghouse (i.e., Protocol Database) is a web-based clearinghouse of sampling protocols used in national parks to monitor the condition of selected natural resources. The database provides a summary of, and in many cases allows the user to download a digital copy of, sampling protocols that have been developed by the prototype monitoring parks or other well-established protocols used in national parks. The Protocol Database also makes it possible to download database components (e.g., tables, queries, data entry forms) in MS Access that are consistent with the Natural Resource Database Template that have been developed for a particular protocol. See the Protocol Database website for available protocols (*http://science.nature.nps.gov/im/monitor/VitalSigns/BrowseProtocol.aspx*).

NR-GIS Data Store
The NR-GIS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR-GIS Data Store is a graphical search interface that links metadata to a searchable data server on which data sets are organized by NPS units, offices and programs. The interface allows customized public or protected searches of natural resource data sets, inventory products and GIS data produced by the I&M and Natural Resource GIS Programs. Each park or network is able to post and curate its data on the server. The NR-GIS Data Store will be integrated with the master NR-GIS Metadata Database application to streamline programmatic data documentation and dissemination processes. The simple browse function of this server can be accessed at: *http://nrdata.nps.gov/*.
See the NR-GIS Data Store website for further information (*http://science.nature.nps.gov/nrdata*).

3.3 – Network Systems Architecture

Rather than developing a single, integrated database system, our design relies upon modular, standalone project databases that share design standards and links to centralized data tables. Individual project databases are developed, maintained, and archived separately. There are numerous advantages to this strategy:

- Data sets are modular, allowing greater flexibility in accommodating the needs of each project area. Individual project databases and protocols can be developed at different rates without a significant cost to data integration. In addition, one project database can be modified without affecting the functionality of other project databases.
- By working up from modular data sets, we avoid a large initial investment in a centralized database and the concomitant difficulties of integrating among project areas with very

different – and often unforeseen – structural requirements. Furthermore, the payoff for this initial investment may not be realized down the road by greater efficiency for interdisciplinary use.

3.3.1 – Project Database Standards

Project database standards are necessary for ensuring compatibility among data sets, which is vital given the often unpredictable ways in which data sets will be aggregated and summarized. When well thought out, standards also help to encourage sound database design and facilitate interpretability of data sets. Databases that are developed for park and network projects will all contain the following main components:

- *Common lookup tables* – These are accessed via links to tables that reside in a centralized database, rather than storing redundant information in each database. These tables typically contain information that is not project-specific (e.g., lists of parks, personnel, and species).
- *Core tables and fields based on network and national templates* – These tables and fields are used to manage the information describing the "who, where and when" of project data. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. These core tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges.
- *Project-specific fields and tables* – The remainder of database objects can be considered project-specific, although there will typically be a large amount of overlap among projects. This is true even among projects that may not seem related – for example, a temperature field will require similar data types and domain values. As much as is possible, efforts will be made to develop these project-specific objects to be compatible with those maintained by other networks and cooperators managing similar data sets – especially if integration with other data sets is important for meeting project objectives.

Compatibility with national standards

As much as possible, NCCN standards for fields, tables and other database objects will mirror those conveyed through the Natural Resource Database Template. Where there are differences between local and national standards, documentation of the rationale for these differences will be developed. In addition, documentation and database tools (e.g., queries that rename or reformat data) will be developed to ensure that data exports for integration are in a format compatible with current national standards.

Centralized database components: common tables

Certain key information is not only common to multiple data sets, but to the organization as a whole – lists of contacts, projects, parks, species are often complex and dynamic. It is a good strategy to centralize this information so that users have access to the most updated versions in a single, known place. Centralizing also avoids redundancy and versioning issues among multiple copies because there is only one copy to maintain. Centralized information is maintained in database tables that can be linked or referred to from several distinct project databases. Network applications – for project tracking, administrative reporting, or budget management – can also

link to the same tables so that all users in the network have instantaneous access to edits made by other users (Figure 3.3).

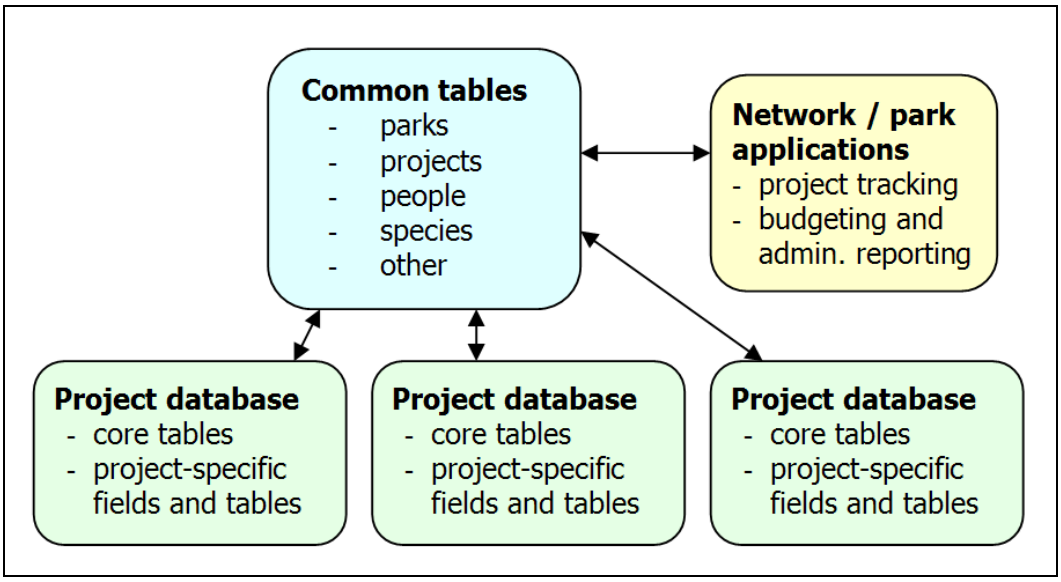

Figure 3.3. Common lookup tables and satellite databases

At present, these common tables are grouped and maintained in separate Access (.mdb) files according to Table 3.1. Separating these tables by functional groupings is done primarily to reduce conflicts and performance losses associated with multiple users in MS Access. Databases associated with individual projects each access the common tables via links established in each project back-end data file. As our network completes its transition to a client-server database system, these components will be placed together in a database that is replicated among data nodes hosted at MORA, OLYM and NOCA. Individual project databases will continue to access these common tables in the same manner.

Table 3.1. Groupings for common lookup tables

Grouping	Description
Parks	list of park units and networks
Projects	list of park and network projects, including inventories, monitoring, park-sponsored initiative projects, and external research projects
People	comprehensive list of contacts for parks and network, project-specific crew lists, lists of groups and users for tracking and managing access privileges
Species	comprehensive list of taxa for the park, linkage to NPSpecies taxonomic module, project-specific species lists
Other Lookups	lists of watersheds, drainages, place names, weather conditions, habitat attributes

Different levels of data standards

The three types of database objects also correspond to three putative levels of data standards. Because common lookup tables are stored in one place and are referred to by multiple databases,

they represent the highest level of data standard because they are implemented identically among data sets. The second level of standards is implied by the core template fields and tables, which are standardized where possible, but project-specific objectives and needs could lead to varied implementations among projects. The third level of standards is applied most flexibly to accommodate the range of needs and possibilities for each project, yet always with compatibility and integrity in mind. The following figure presents the resulting variation in implementation of these differing levels as a "bull's eye", with the common lookup tables providing the most consistent implementation and hence the smallest range of variation.

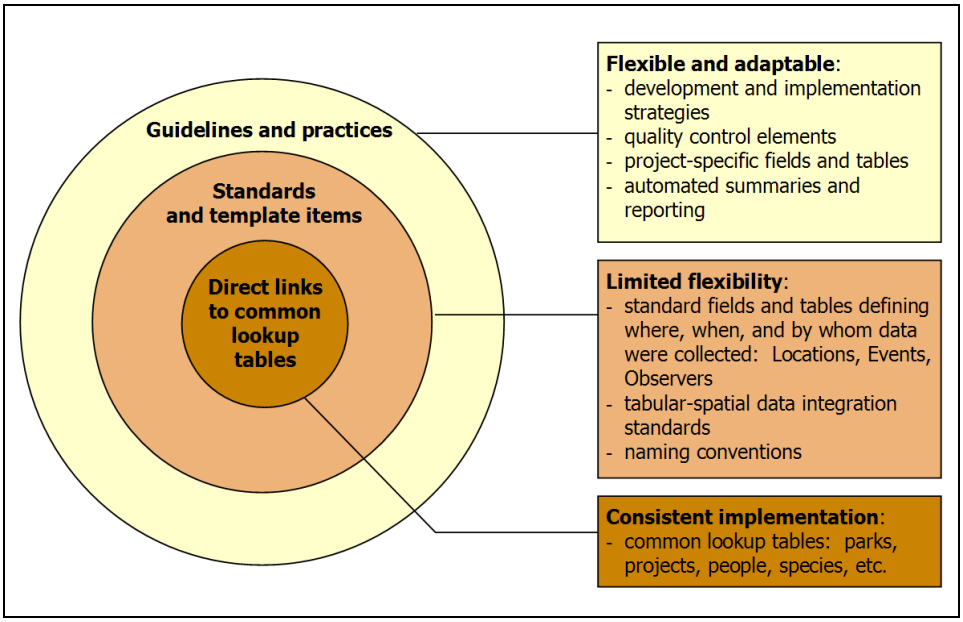

Figure 3.4. Different levels of data standards and their corresponding degree of implementation variability

3.3.2 – Client-Server Database Architecture

Our data management program is in the process of completing a transition to a client-server database system. Our strategy is to manage common tables and high-value, long-term project databases within this system as a means of maximizing performance in a distributed, multi-user environment. There are three data servers that comprise the network infrastructure – one located at MORA, and one at each of the prototype parks (NOCA and OLYM). These three servers function as independent data nodes that can be accessed from any park location so long as it is within the wide area network maintained by the Pacific West Region. They are also integrated in that common tables are replicated regularly among data nodes, backups for one node are stored on a separate node, and network databases are distributed across the three nodes.

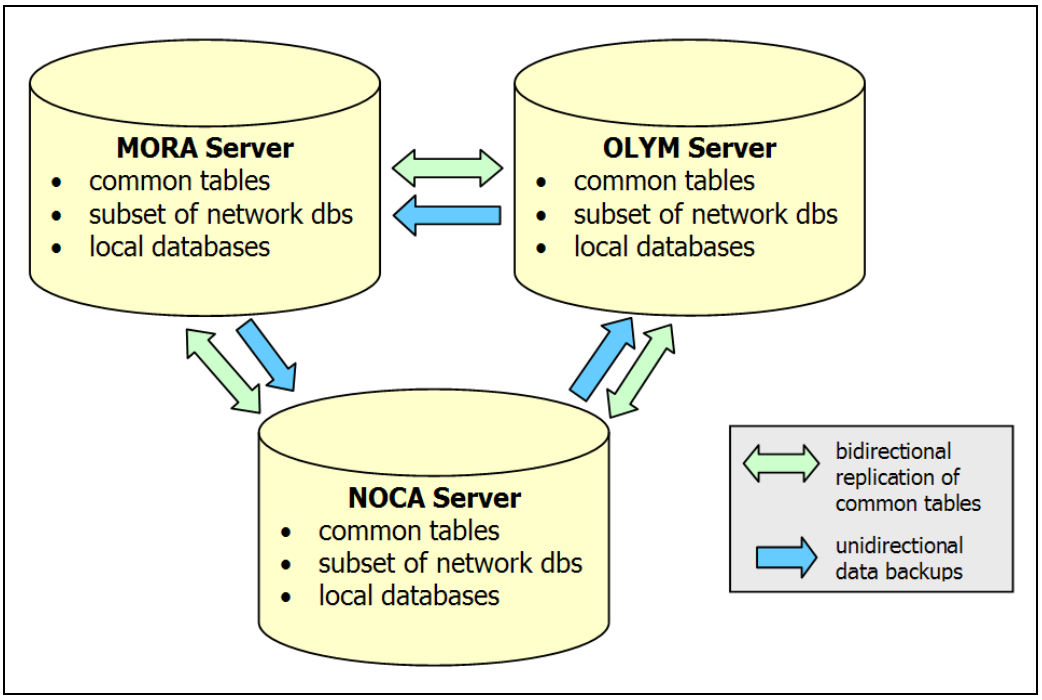

Figure 3.5. Client-server database architecture for NCCN

The advantages of this strategy include:

- Data are fully backed up on an off-site data node, which is crucial for information recovery in case of a local catastrophe at one of the host sites. Backups can be automated through scheduled services.
- Common tables are replicated for continuous access during server down time. Also, by hosting common tables at each node, locally hosted databases can refer to the local copy of the common tables, which maximizes performance and server node autonomy.
- Parallelism in data node setup simplifies system maintenance, which is important given our collaborative model for data management.
- Distribution of data services allows for program growth and balanced access loads.
- Park databases can be hosted and managed locally.

3.3.3 – Project Tracking Application

To support program coordination and annual reporting, and to improve accountability for the products of our natural resource inventory and monitoring efforts, our network will develop and implement a project tracking database. The primary functions of this application include:

- Maintaining the list of projects – By maintaining a single list of natural resource data projects, it is much easier to quickly find project-related information (e.g., status, funding sources and amounts, objectives, contact information) and summarize that information for administrative reports.
- Tracking product deliverables – For each project a comprehensive list is maintained of what deliverables are expected and when. Once they are delivered and posted or archived, this

function shifts to being a search tool (finding aid) for available products. Deliverables are first specified at project initiation and information is updated at various project milestones (e.g., contracting, product delivery, archival).

- Managing project codes – These are intelligent alphanumeric codes used to tie together digital information in various, minimally connected systems (e.g., RPRS, PMIS), along with analog materials that cannot otherwise be linked to an integrated information system. These codes are also used to link to databases and GIS themes, especially where information from multiple sources is stored together.

This application will be hosted by each network data node (Figure 3.4), and will be replicated among nodes to ensure uninterrupted access to NCCN users. Although primarily maintained by NCCN data managers, the database will be available to project leaders, GIS staff, the network coordinator, and NCCN park administrators. Each of these staff will be able to make certain changes to update information about project status, deliverable details, etc. Certain database views will be created to help project leaders keep on schedule and to facilitate quick reporting on project status, accomplishments and delivered products.

This section will be updated once the application is implemented and additional details are available.

Credits

Section 3.2 (National Information Management Systems) was adapted from material written by Lisa Nelson (WASO). Thanks to Patrick Flaherty for ideas and material related to Section 3.1 (Computer Resources).

Chapter 4 – Data Management Process and Work Flow

This chapter considers the general work flow characteristics of projects that produce natural resource data and then gives an overview of how natural resource data are generated, processed, finalized and made available. Data management activities that relate to the various stages of a project are highlighted. By describing the progressive stages of a project and the life cycle of the resulting data, we can more easily communicate the overall objectives and specific steps of the data management process. In addition, this awareness helps us to manage the staffing resources needed to produce, maintain, and deliver quality data and information. More details about data acquisition, quality assurance, documentation, dissemination and maintenance can be found in later chapters of this plan.

4.1 – Project Work Flow

From the perspective of managing workflow, there are two main types of projects:

- *Short-term projects*, which may include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring.
- *Long-term projects*, which will mainly be the implemented monitoring projects central to the I&M program but which may also include multi-year research projects and monitoring performed by other agencies and cooperators. Long-term projects will often require a higher level of documentation, peer review and program support.

From a data management standpoint, a primary difference between short- and long-term projects is an increased need to adhere to standards for long-term projects to ensure internal consistency over time. While the need to follow standards is still present for short-term projects, sometimes the cost of compliance will outweigh the benefits due to the scope, budget, and level of NPS influence over the project details. Nevertheless, both short-term and long-term projects share many work flow characteristics, and both generate data products that must be managed and made available.

Projects can be divided into five primary stages: planning and approval; design and testing; implementation; product integration; evaluation and closure (Figure 4.1). Each stage is characterized by a set of activities carried out by staff involved in the project. Primary responsibility for these activities rests with different individuals according to the different phases of a project. Additional discussion of the different roles and responsibilities of park and network staff can be found in Chapter 2 of this plan.

4.1.1 – Planning and Approval

Here many of the preliminary decisions are made regarding project scope and objectives. In addition, funding sources, permits and compliance are all addressed in this phase. Primary responsibility rests with project leaders and program administrators. Although this phase lacks specific data management activities, it is important that data managers remain informed of projects in this phase. This is especially true as timelines for deliverables are finalized. All contracts, agreements and permits should include standard language that describes the formats, specifications, and timelines for project deliverables.

4.1.2 – Design and Testing

During this phase, all of the details are worked out regarding how data will be acquired, processed, analyzed, reported and made available to others. The project leader is responsible for developing and testing project methodology, or for modifying existing methods to meet project objectives. It is critical that the project leader and the data manager work together throughout this phase. The dialog between these two will help to build and reinforce good data management throughout the project – especially during the crucial stages of data acquisition, processing, and retrieval. By beginning collaborative development as soon after project approval as possible, data integrity and quality can most easily be assured. An important part of this collaboration is the development of the data design and data dictionary, where the specifics of database implementation and parameters that will be collected are defined in detail. Devoting adequate attention to this aspect of project is possibly the single most important part of assuring the quality, integrity and usability of the resulting data. Although it is likely that additional changes will need to be made once implementation has begun, it is important to minimize these post-hoc changes through careful design work. Once the project methods, data design, and data dictionary have been developed and documented, a database can be constructed to meet project requirements.

4.1.3 – Implementation

During the implementation phase, data are acquired, processed, error-checked and documented. This is also when products such as reports, maps, GIS themes, and other products are developed and delivered. The project leader oversees all aspects of implementation – from logistics planning, contracting, training, and equipment procurement to data acquisition, report preparation and final delivery. Throughout this phase, data management staff function primarily as facilitators – providing training and support for database applications, GIS, GPS and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development. The specific roles of data management staff during this phase will depend primarily on the technical capabilities of the project staff. As much as is possible, these roles should be worked out in advance of implementation.

Toward the end of this phase, project staff members work to develop and finalize the deliverables that were identified in the project planning documents (i.e., protocol, study plan, contract, agreement or permit). In general, all raw and derived data products, metadata, reports and other documentation should be delivered to the data steward assigned to the project. Administrative records should be delivered to appropriate park and network staff as specified. All project deliverables should be developed and delivered according to product specifications, which should be stipulated in all protocols, contracts, agreements, and permits. Products that do not meet program requirements will be returned for revision.

4.1.4 – Product Integration

During this phase, data products and other deliverables are integrated into national and network databases, metadata records are finalized and posted in clearinghouses, and products are distributed or otherwise made available to their intended audience. Another aspect of integration is merging data from a working database to a master database maintained on the network server. This occurs only after the annual working data set has been certified for quality by the project leader. Certain projects may also have additional integration needs, such as when working jointly with other agencies for a common database.

Product integration includes creating records for reports and other project documents in NatureBib, posting imaged documents to the appropriate repository, posting metadata records that have been completed and submitted by project leaders, and updating NPSpecies to reflect any new species occurrence information derived from the project. This will allow the information from the project to be searchable and available to others via service-wide search engines.

4.1.5 – Evaluation and Closure

Upon project closure, records are updated to reflect the status of the project and its associated deliverables in a network project tracking application. For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season, and leads to an annual review of the project. For non-cyclic projects, this phase represents the completion of the project. After products are catalogued and made available, program administrators, project leaders, and data managers should work together to assess how well the project met its objectives, and to determine what might be done to improve various aspects of the methodology, implementation, and formats of the resulting information. For monitoring protocols, careful documentation of all changes is required. Changes to methods, SOPs and other procedures are maintained in a tracking table associated with each document. Major revisions may require additional peer review.

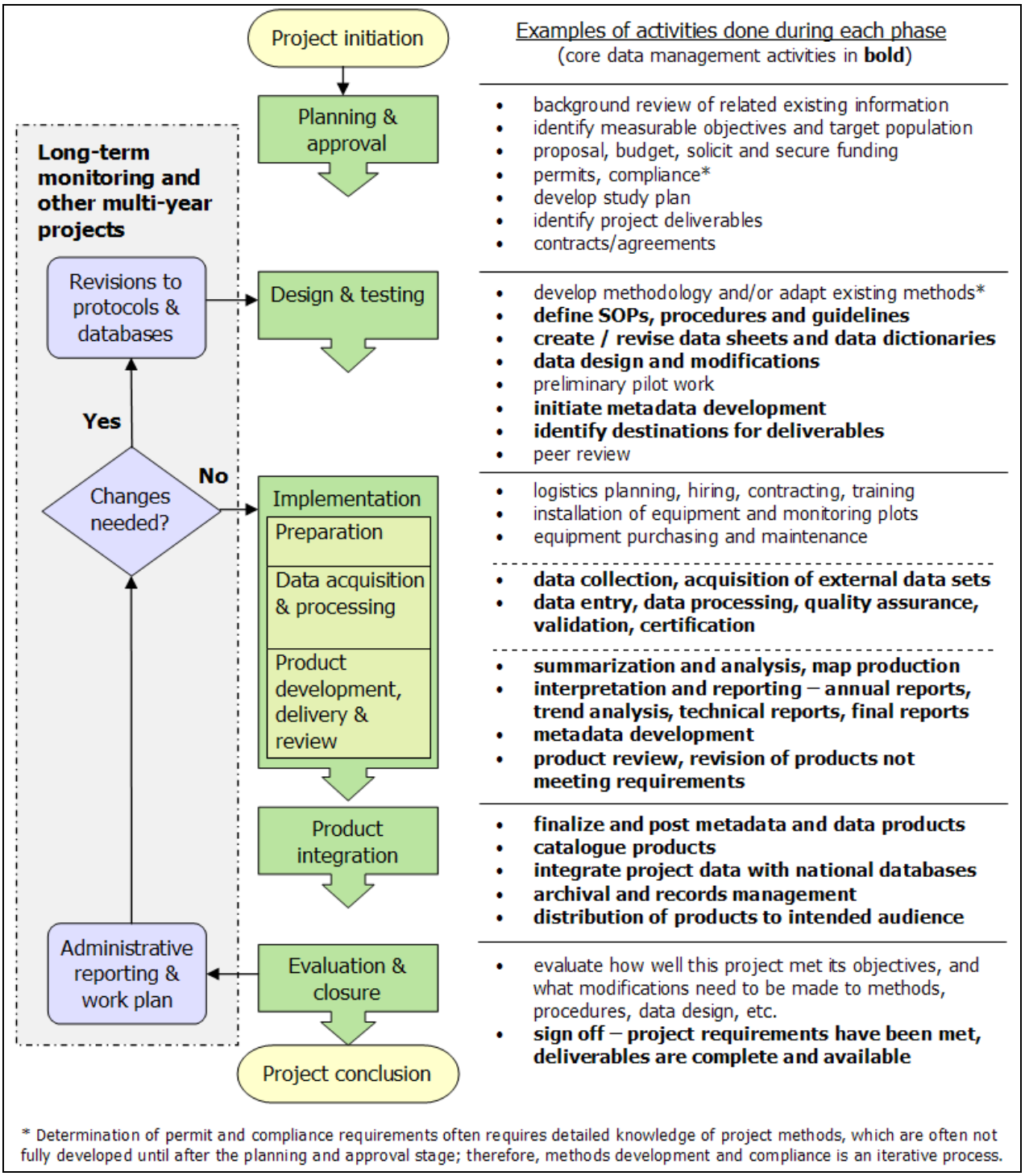

Figure 4.1. Conceptual model of project work flow

4.2 – Data Life Cycle

During various phases of a project, project data take on different forms and are maintained in different places as they are acquired, processed, documented and archived. This data life cycle is

characterized by a sequence of events that we can model to facilitate communication. These events involve interactions with the following objects:

- *Raw data* – Analog data recorded by hand on hard-copy forms and digital files from handheld computers, GPS receivers, telemetry data loggers, etc.
- *Working database* – A project-specific database for entering and processing data for the current season (or other logical period of time). This might be the only database for short-term projects where there is no need to distinguish working data for the current season from the full set of validated project data.
- *Certified data and metadata* – Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and ready for distribution. Metadata records include the detailed information about project data needed for proper use and interpretation (see Chapter 7).
- *Master database* – Project-specific database for storing the full project data set, used for viewing, summarizing, and analysis. Only used to store data that have passed all quality assurance steps.
- *Reports and data products* – Information that is derived from certified project data.
- *Edit log* – A means of tracking changes to certified data.
- *National databases and repositories* – Applications and repositories maintained at the national level, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public.
- *Local archives and digital library* – Local storage of copies of data, metadata and other products generated by projects. Archives are for hard-copy items and off-line storage media, whereas the digital library is maintained live on a server.

Although the data life cycle may vary depending on specific project needs and objectives, the typical life cycle for NCCN proceeds as follows (Figure 4.2):

1. *Acquire data* – For data recorded by hand in the field, data forms should be reviewed regularly (at least daily) for completeness and validity in order to capture errors as close to their origin as possible.

2. *Archive raw data* – Copies of all raw data files are archived intact. Digital files are copied to the digital library section for the project; hard copy forms are either scanned and placed in the digital library or are copied and placed in the archives. Archiving or scanning of hard copy data forms may occur at the end of a season as a means of retaining all marks and edits made during the verification and validation steps.

3. *Data entry / import* – Analog data are entered manually and digital data files are uploaded to the working database.

4. *Verification, processing and validation* – Verify accurate transcription of raw data; process data to remove missing values and other data flaws; and validate data using database queries to capture missing data, out-of-range values, and logical errors.

5. *Documentation and certification* – Develop or update project metadata and certify the data set. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and documented. It also means that data and metadata are ready to be posted and delivered.

6. *Archive versioned data set* – Copies of the certified data and metadata are placed in the digital library. This can be accomplished by storing a compressed copy of the working database or by exporting data to a more software-independent format (e.g., ASCII text; see Chapter 10).

7. *Extract data, post data, and update national databases* – To make data available to others, certified data and metadata are posted to national repositories such as NR-GIS Data Store. In addition, national databases such as NPSpecies, NPSTORET, and NR-GIS Metadata Database are updated with data extracted from the certified data set. Note: Data and data products may not be posted if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern (see Chapter 9).

8. *Upload data* – Certified data are uploaded from the working database to the master project database. This step might be skipped for short-term projects where there is no need to distinguish working data for the current season from the full set of certified project data.

9. *Reporting and analysis* – Certified data are used to generate data products, analyses, and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data might be exported for analysis or summarized within the database.

10. *Store products* – Reports and other data products are stored according to format and likely demand – either in the digital library, on off-line media, or in the document archives.

11. *Post products and update national databases* – To make data available to others, reports and other products are posted to national repositories such as NR-GIS Data Store or the NR Data Image Server. In addition, products are catalogued in NatureBib. Data products may not be posted if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern (see Chapter 9).

12. *Distribute data and information* – Data, metadata, reports and products can be shared and distributed in a variety of ways – especially via the web-based national databases and repositories, by FTP or mailing in response to specific requests, or by providing direct access to project records to cooperators. In all cases, distribution will follow legal requirements under the Freedom of Information Act and limitations established to protect information about sensitive resources (see Chapter 9).

13. *Track changes* – All subsequent changes to certified data are documented in an edit log, which accompanies project data and metadata upon distribution. Significant edits will trigger reposting of the data and products to national databases and repositories.

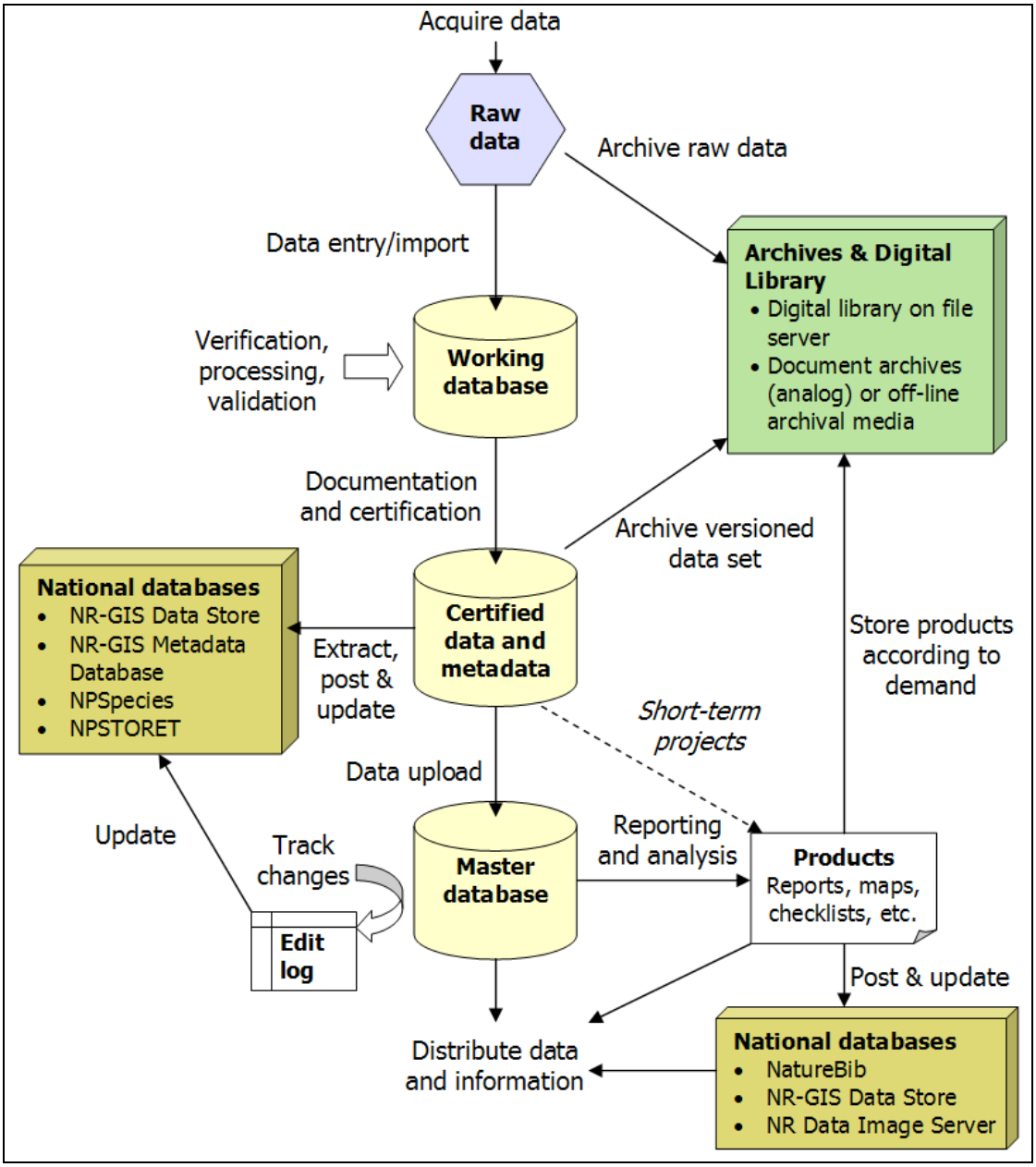

Figure 4.2. Diagram of the typical project data life cycle

This sequence of events occurs in an iterative fashion for long-term monitoring projects, whereas this sequence is followed only once for short-term projects. For projects spanning multiple years, decision points include whether or not a separate working database is desirable, and the extent to which product development and delivery is repeated year after year.

4.3 – Integrating and Sharing Data Products

Once project data and data products have been finalized, they need to be secured in long-term storage and made available to others. To accomplish this requires that we use a range of information systems such as product repositories, clearinghouses, and web applications. Each of these systems has a different purpose and function, as shown in Figure 4.3.

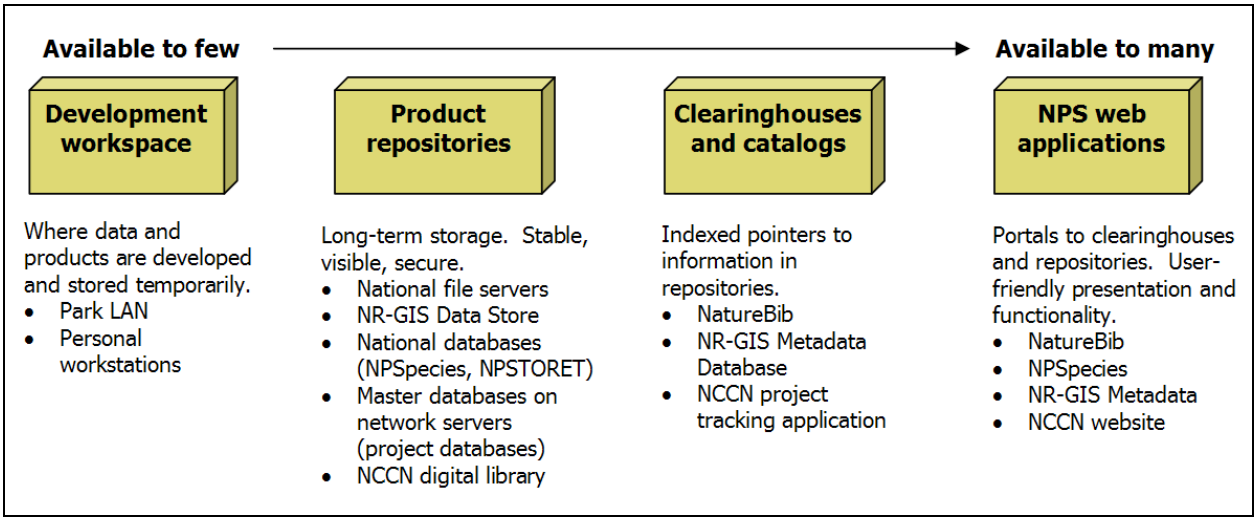

Figure 4.3. Storing and disseminating project information

The specific repositories for most NCCN products are indicated in Table 4.1.

Table 4.1. Repositories for NCCN products

Repository	Item
NCCN Digital Library	Project data, metadata, and other products Raw and certified data sets Metadata, protocols, SOPs Reports and administrative records Digital photographs, derived products
NCCN Project Databases	Comprehensive data for multi-year projects
Park Collections and/or National Archives	Administrative records, voucher specimens*, raw data forms, hard copy reports
National Databases - NPSTORET, NPSpecies, NatureBib	Compiled information about water quality, park species lists and taxonomic documentation, park resource bibliographies
NR Data Image Server	Copies of digital reports and other documents (catalogued in NatureBib)
NR-GIS Data Store	Copies of digital data sets (non-sensitive) and metadata

* Biological specimens can also be retained at other facilities (e.g., University of Washington Herbarium) with an appropriate agreement.

4.3.1 – Data Distribution

The process of product distribution involves several steps (Figure 4.4). As products are finalized, they can be sent to the appropriate person for integration, posting and distribution. In most cases it will be either the Data Manager or GIS specialist who reviews the product for conformance with format standards, then stores the product in the appropriate repository. Note that it is expected that all products will have already been reviewed for completeness and accuracy prior to delivery. After storing the products, their existence is documented by posting metadata and by updating records in the NCCN project tracking application. At this point, data discovery is accomplished as metadata are then indexed by the clearinghouse function of the NR-GIS Metadata Database. These metadata records provide pointers to data and data products. Distribution then follows as data discovery allows potential users to find and either request or download the data sets from their repositories.

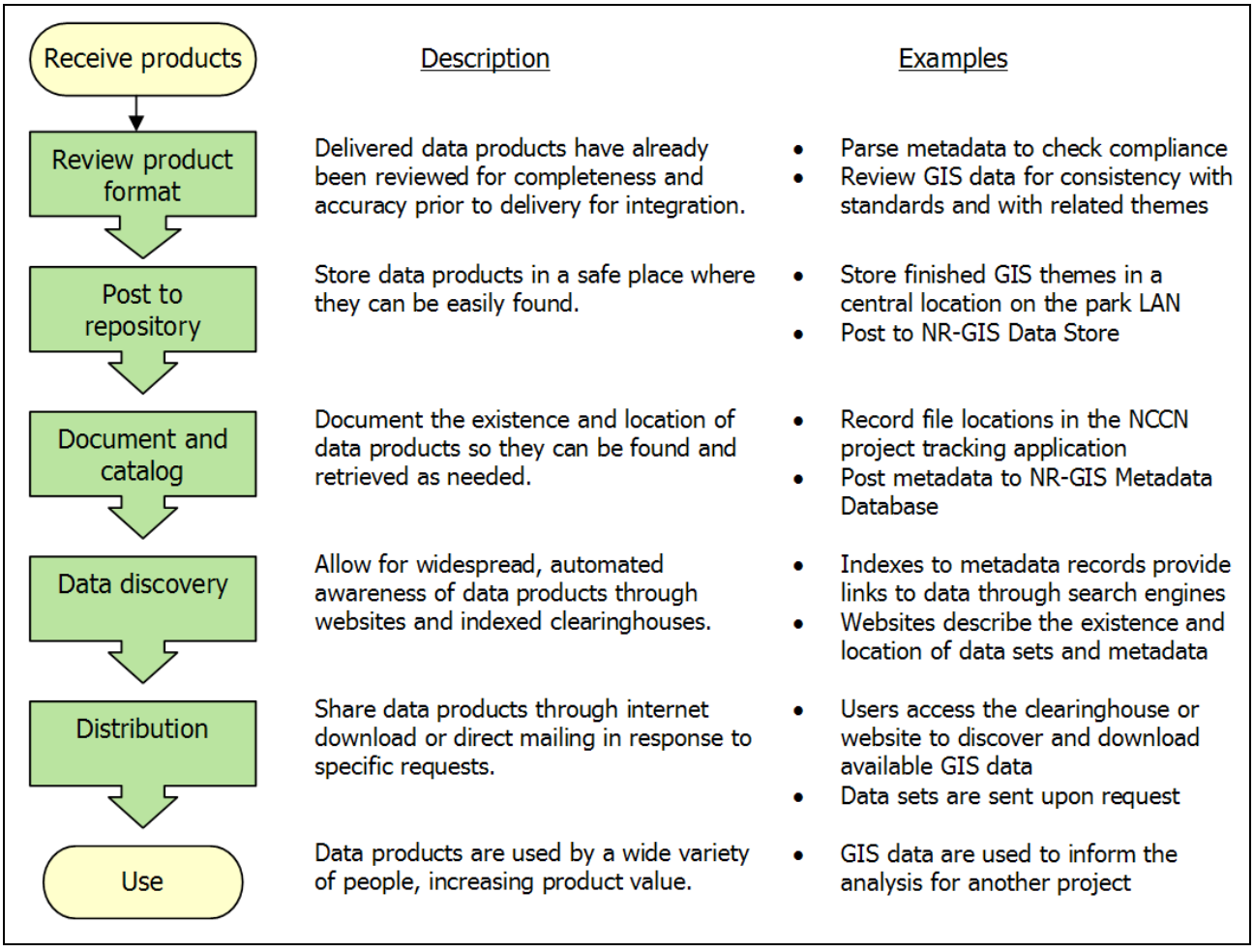

Figure 4.4. Steps involved in product distribution

4.3.2 – Integration with National Databases

In addition to storing and distributing data products, product integration also involves updates to national databases such as NPSpecies and NPSTORET. Both of these databases have local desktop databases which can be updated with data collected during the course of a project. Desktop databases are then uploaded and synchronized with the national databases on a regular basis.

To update NPSpecies, data on the distribution and occurrence of species in NCCN parks will be compiled and added to the database upon delivery of data and data products. Synchronization with the master version of NPSpecies will occur at least twice annually, or more frequently depending on the timing and amount of updates.

For NPSTORET, any project collecting water quality data will be flagged in the project tracking application so that water quality data can be either extracted and uploaded or directly entered into NPSTORET. All water quality data collected by our network will be managed according to guidelines from the NPS Water Resources Division. We will implement and maintain a desktop copy of NPSTORET and transfer its contents at least annually to NPS Water Resource Division for upload to the STORET database (Figure 4.5).

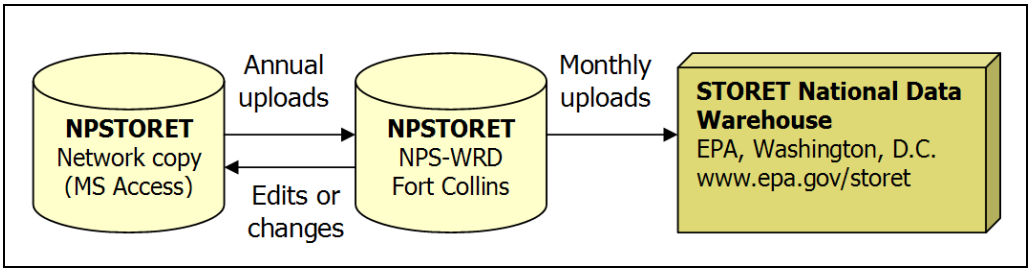

Figure 4.5. Data flow diagram for water quality data

Credits

This chapter was adapted from concepts and material developed in collaboration with Dorothy Mortenson (Southwest Alaska Network), Velma Potash (Cape Code National Seashore), Sara Stevens (Northeast Coastal and Barrier Network), and Doug Wilder (Central Alaska Network). Special thanks to Dorothy Mortenson and Doug Wilder for significant contributions and ideas.

Chapter 5 – Data Acquisition and Processing

The NPS I&M Program, in support of the Natural Resource Challenge, is responsible for acquiring the necessary information required by park managers to properly manage and maintain the natural resources of their park. To successfully accomplish this task, information from multiple sources is collected and appropriately processed so that it meets NPS data standards. This chapter of the Data Management Plan describes the steps involved with acquiring data as well as the initial stages of data processing. These steps are integral to strengthening the scientific foundation of the I&M program and providing the park managers with high quality data necessary to properly manage park resources.

There are three general classifications for the types of data handled by the I&M Program:

- **Programmatic Data** – any data produced from projects that are initiated (funded) by the I&M Program or projects that in some way involve the I&M Program.

- **Non-programmatic NPS Data** – any data produced by the NPS that did not involve the inventory and monitoring program.

- **Non-programmatic External Data** – any data produced by agencies or institutions other than the National Parks Service.

These definitions do not in any way indicate or rank the importance of the different data types to the I&M Program. This is only a distinction between data gathered from different sources. The importance or value placed on a data set will be based on the quality or usefulness of data itself and the impact it has on the NCCN I&M Program and parks.

The following sections outline in more detail the manner in which the different data types described above are acquired and processed.

5.1 – Programmatic Data

Data collection initiated by the Inventory and Monitoring Program usually takes place as part of either natural resources inventories or vital signs monitoring projects but can also include collaborative efforts between the NCCN I&M Program and other regional programs or institutions. Data for these projects are typically collected by I&M personnel, park staff or by cooperators/contractors.

Natural Resources Inventories

Resource inventories were designed to identify the primary resources of each park and as such they represent an important phase in the management of park resources.

Vital Signs Monitoring Projects

Vital signs are measurable, early warning signals that indicate changes that could impair the long-term health of natural systems. Early detection of potential problems allows park managers to take steps to restore ecological health of park resources before serious damage can happen. Each Network, working in conjunction with park resource managers, was given the task of prioritizing the natural resources vital signs that were most important to its parks and to the network as a whole.

5.1.1 – Data Discovery and Data Mining

Data discovery or data mining is the process of searching for existing data/information that may be useful to the I&M Program mission and that is related to the natural resources of the network parks. For more information on I&M data mining practices refer to:

http://science.nature.nps.gov/im/inventory/index.cfm

A large percentage of data discovery occurs at the onset of new projects or during the development of new protocols. The process involves reviewing many different sources for varying types of information. Many of the following data sources are accessible via the internet, but some require visiting local research or academic institutions, museums or local parks to conduct the searches.

Bibliographic/Literature

- National NPS Databases (e.g. NatureBib)
- Online literature databases (e.g. First Search or Biosis)
- Local document library (e.g. NCCN Library)
- Library catalogs (e.g. academic or research institutions)
- Park archives

Geographic Data

- Regional GIS Specialists
- Park GIS Specialists
- Federal and State Geographic data clearinghouses

Biological/Natural Resources Data

- NPSpecies
- Voucher collections (museums, parks, universities)
- Network Parks

All information collected during the data discovery process is maintained either electronically or in hard copy format, depending on how it was collected. Any geographic data sets collected during this process should be accompanied by FGDC compliant metadata, and all data sets found – geographic or otherwise – should be entered in Dataset Catalog.

- **Dataset Catalog** – new data sets (spatial or non-spatial) should be entered into Dataset Catalog.

- **NatureBib** – the desktop version of NatureBib has been adopted by the NCCN to function as the internal literature catalog for the Network. All natural resource reference information collected during the data discovery process is entered into NatureBib.

- **NPSpecies/NatureBib** – Information relating to the biodiversity of Network parks is entered into NPSpecies and linked to the associated reference (assuming the data came from a report or publication) in NatureBib.

- **Reference Cabinets** – Hard copy materials are stored in three locations depending on how they relate to the Network parks and the NCCN I&M Program.

All original data sheets, final reports, and contracts are stored in a fire proof cabinet.
General references as well as those references linked to NPSpecies records in NatureBib are stored in file cabinets in the NCCN I&M Office.

Much of the information collected during the data mining process is likely to be *legacy data*, or data collected prior to the inception of the NCCN I&M Program. If legacy data are collected in a digital format the information should be converted to current file formats compatible with the current software standards. In the past, hard copy materials were maintained as such, but in the future, the NCCN intends to scan hardcopy references and materials and saving them as .pdf files in order to create a digital library.

Data discovery is an integral part of project development, but data discovery efforts should not be limited solely to project development needs. This should be an ongoing process requiring regular data searches and visits to Network parks to ensure that the NCCN I&M Program maintains as much relevant material pertaining to the parks as possible. Encouraging data sharing among Network parks will assist in this process and may alleviate the need for regular searches of park records. Our goal is to conduct general data discovery efforts on an annual basis. (*Guidelines in development*)

5.1.2 – Field Studies

Biological inventories and monitoring projects are the most common examples of field studies conducted by the NCCN I&M Program. The Data Manager is responsible for ensuring that data collection, entry, verification, storage, and archiving for all field projects are consistent with the NCCN I&M Program standards. In addition to general standard operating procedures (SOPs) that define network-wide requirements, protocol specific SOPs are developed detailing procedures and/or methodologies. The data manager will work closely with the principal investigator and Network staff to develop these guidelines and methodologies connected with data collection, storage and maintenance. This may range from detailing the proper usage of data entry forms or databases to outlining calibration procedures for automated data loggers. Refer to the individual protocols for protocol specific SOPs. For general data management guidance, refer to:

http://science.nature.nps.gov/im/units/nccn/datamgmt_guide.cfm

Field Data Acquisition Methods

Listed below are some of the tools available for field data collection. Refer to the protocol specific SOPs for details on how the following tools may apply to individual projects.

Field Forms – the most common method of recording field data. Inexpensive but more opportunities for errors during the collection/data entry process. Requires neat, legible handwriting and rigorous QA/QC.

Field Computers – increase data collection and data entry efficiency. Data can be directly dumped from the field computers to the office desktops thereby eliminating the data entry step. Fewer chances for error as QA/QC checks can be built into the database. Could be inefficient if copious amounts of notes or comments need to be recorded in the field.

- **Palm-top computers (PDAs) –** the small size and relative low cost of these devices make them attractive options for collecting field data. Good for small field projects but not powerful enough for large, data intensive field projects. PDAs can be weatherized fairly easily and inexpensively. Most run either Windows CE or Palm operating systems which may require additional processing/programming to transfer/create the database structure in the field units.

- **Tablet PCs –** same properties as most laptops and provide the user with the convenience of a touch screen interface. They are bulkier, more expensive and harder to weatherize than the PDAs but are more powerful as well. Good for field projects that are very data intensive. Because these units run Windows XP (Tablet Edition), the project database can be directly transferred from desktop units to field units without additional programming steps.

Automated Data Loggers – mainly used to collect ambient information such as weather data or water quality information. Must be properly calibrated and maintained so field crews must receive proper training and SOPs outlining these procedures.

- **Permanently deployed devices –** often cost prohibitive. Data from these devices must be retrieved and batteries changed on a regular basis. These intervals should be defined in the protocol.

- **Portable hand-held devices –** deployed for sampling only during site visits. Generally less expensive than units that are permanently deployed in the field.

GPS Units – Two types of GPS units are often used during field work in NCCN parks to collect location information. (*Guidelines in development*)

- **Handheld Garmin Units –** good for collecting general position information. Not recommended for obtaining high accuracy location information.

- **Trimble and GeoExplorer GPS Receivers –** good for collecting highly accurate (sub-meter) location information. Due to the small size of certain of the NCCN parks, field crews are encouraged to use these units when collecting GPS data.

5.1.3 – Data entry

All field studies funded by the NCCN I&M Program have a Microsoft Access database associated with them. The NCCN I&M Program has adopted the Natural Resources Database Template (NRDT) as the foundation for its databases and will provide assistance in database development for any programmatic project. The database template is highly flexible and can be modified and customized for each project to meet the needs and requirements of the researcher. The database incorporates mechanisms such pick lists and validation rules for quality assurance purposes. (*Guidelines in development*)

- Field crews/project staff enter all data into the database provided to them.
- Field crews are required to periodically forward the data manager the project data files. Refer to individual protocols for the requirements regarding forwarding data.
- The data manager maintains the master copy of the database and updates it with data files received from the field crews.
- All data must undergo QA/QC procedures (see Chapter 6 of this document for more specifics relating to data verification and validation).

5.1.4 – NPSpecies and Biodiversity Data

NPSpecies is a National Park Service database developed by WASO to store, manage and disseminate scientific information on the biodiversity of organisms in National Park Service units throughout the United States and its territories. The database is available in an on-line form (Oracle) or a desktop version (MS Access). For more specifics on NPSpecies please refer to the following web page:

http://science.nature.nps.gov/im/apps/npspp/index.cfm

Three forms of evidence used to document a species existence in NPS units are:

- *References*
    - All written reports which document species presence in a park unit are entered into the NPSpecies and NatureBib databases.

- *Observations*
    - All observation data collected during biological inventories are entered into the NPSpecies database.

- *Vouchers*

- Information relating to all voucher specimens collected during a biological inventory must be entered into NPSpecies.
- Once catalog and accession numbers are obtained, all voucher information must also be entered into ANCS+.
- The parks where the vouchers were collected technically own the specimens and have the right to decide where the specimens will be stored.

5.1.5 – Changes to Data Collection Procedures/Protocols

Changes to established data collection procedures are discouraged unless there are acceptable, valid reasons for altering the methodologies. Ideally, all problems should be identified during the design and testing stages of the project and changes implemented prior to the collection of any field data. Protocols should attempt to identify any foreseeable issues that might occur as well as contingencies to address them. Inevitably, unforeseen problems may occur which require procedures/protocols revision after data collection has begun. Significant changes to the protocols must be approved by the principal investigator, project leader and data manager. The key official must evaluate the proposed changes and determine if additional peer review is required before accepting them.

Altering data collection procedures or protocols may also occur as a result of the comprehensive review that all monitoring protocols undergo every five years. During the review, data are evaluated to determine if the current protocol accomplished its goal. If it is concluded that the protocol in its present form has not achieved the desired results, changes could be recommended. Once again, all changes must be approved by the principal investigator as well as the key official and data manager.

5.2 – Other NPS Data

A large percentage of data collected in Network parks are collected by the park personnel involved in projects initiated at the individual park level or by other NPS regional or national programs. The data collected and products produced by such efforts provide a great deal of information about park natural resources and are therefore relevant to the mission of the I&M Program.

5.2.1 – Park Data

Parks in the NCCN often use base funding or receive funding through NRPP (Natural Resources Protection and Preservation) programs to support park-level projects.

- **Park-based biological inventories** - Network parks often conduct their own park-based inventory projects, the data from which can be used to supplement Network-level inventories conducted by the I&M Program.
- **Park-based monitoring projects** - Parks also engage in park-level monitoring projects (such as vegetation and water quality) which produce information that is valuable when developing Network-level monitoring protocols.

- **Park and multi-park based projects** - other studies or projects conducted at the park or regional level that do not fall into one of the previous two categories (e.g. restoration projects).

5.2.2 – Regional and National Programs

NPS regional and national programs support all of the parks within the Pacific West Region and are good resources for natural resources information. Table 5.1 lists the regional contact information for the personnel in charge of the following programs in the NCCN.

- **Air –** National-scale programs collect data, maintain databases, assure data quality, and perform the trend analyses relevant to NCCN air quality issues. The NCCN I&M program will rely on the data analyses from these national scale monitoring networks to obtain trends for many of its air vital signs.
- **EPMT -** Exotic Plant Management Teams (EPMT) collect and maintain data regarding the presence of exotic species in regional parks, and develop and document the methods used to treat these species. This information is stored in the Alien Plant Control and Monitoring Database (APCAM) which is maintained by the EPMT data manager or EPMT liaison. http://www1.nrintra.nps.gov/brmd/invasivespecies/exoticplants/index.cfm
- **Fire Program – D**ata concerning the occurrence of fires within the NCCN are maintained both at the individual park and regional level. National databases such as Fire-Pro, SACS and the soon-to-be implemented Fire Program Analysis (FPA) package (http://fpa.nifc.gov/) have been and will be used to maintain information regarding fire incidence and the resources dedicated to fire management. The NPS is also involved in efforts such as the Joint Fire Science Program (http://www.firescience.gov/) that provide scientific information and support for fuel and fire management programs.
- **GIS –** The NCCN I&M Program is supported by regional GIS specialists to help ensure that regional GIS data are available and accurate. Much of this data is also available through the Spatial Data Clearinghouse.
- **Inventory and Monitoring –** The regional and network coordinators provide support and guidance for I&M related tasks and activities and communicate national program directives.
- **Water -** Almost all of the field data collection by the regional water resources program is done in support of the water resources vital signs monitoring projects. The program also synthesizes, analyzes and interprets water resources data collected by parks.
- **Wildlife Management -** The regional Threatened and Endangered (T&E) Species Specialist provides management support to the parks in the Pacific West Region. The T&E Species Specialist coordinates reporting of T&E species populations.

Table 5.1. NPS Pacific West Region program contacts

Program	Contact	Title	E-Mail
Air	Judy Rocchio Elizabeth Waddell	PWR Air Resource Lead Air Resource Specialist	judy_rocchio@nps.gov elizabeth_waddell@nps.gov
EPMT	Judy Daniels Todd Neel	EPMT Data Manager NCCN EPMT Coordinator	judy_daniels@nps.gov todd_neel@nps.gov
Fire	Jeff Manley	Fire Pro Core Team	jeff_manley@nps.gov
GIS	Craig Dalby Allen McCoy	Regional GIS Coordinator Regional GIS Specialist	craig_dalby@nps.gov allen_mccoy@nps.gov
Inventory & Monitoring	Penny Latham Mark Huff	PWR Reg. Coordinator NCCN Coordinator	penny_latham@nps.gov mark_huff@nps.gov
Water	Mietek Kolipinski Marie Denn	PWR Water Res. Lead Regional Aquatic Ecol.	mietek_kolipinski@nps.gov marie_denn@nps.gov
Wildlife	Steve Gibbons	T & E Specialist	steve_gibbons@nps.gov

Data Processing

These data often do not require a great deal of processing because the NCCN I&M Program shares many of the file standards with Network parks and regional programs. Some basic processing steps include:

- Enter all new park biodiversity data into NPSpecies (this is especially important for park-based biological inventories) and enter all associated references into NatureBib.
- Enter all natural resource reports and publications related to NCCN parks into NatureBib. Hard copies should be stored in the appropriate park collections and electronic copies archived in the proper directory on the NCCN file servers.
- Ensure that all GIS data are in the proper projection and accompanied by FGDC-compliant metadata.
- All data sets should be entered into and tracked using Dataset Catalog.

It is important that park, regional and network staff work closely together to ensure that information is maintained in a manner that promotes data sharing. Accordingly, the NCCN Data Manager will:

- Work closely with park and regional personnel to ensure that high quality data are available
- Provide training to park staff interested in learning to use NPSpecies or NatureBib
- Offer to enter any information collected/maintained by the parks into NPSpecies or NatureBib
- Provide assistance with the processing of voucher specimens collected in Network parks
- Develop databases based on the NRDT that meet the needs of park resource managers

5.3 – External Data

As was the case in the previous section, external data sources often provide information relevant to the mission of the I&M program. It should be noted that such sources need not be directly

connected to NCCN parks but may instead pertain to methodologies or protocols that could assist NCCN staff with the development of a more productive program.

There are a myriad of possible data sources available to provide information pertinent to the NCCN I&M Program.

- Government Agencies - Federal, State and local
- Academia
- Private organizations/non-profit groups

Some NPS units in the NCCN collect these data and contribute them to the larger regional or national data sets maintained by the other agencies. The agencies or organizations that compile these data have the expertise to conduct the proper quality control procedures and the capability to function as a repository and clearinghouse for the validated data.

When the data are not kept in-house, there can be two modes of acquisition: downloads from online databases or requests for data on CD or other media. In some cases, portions of external databases may be incorporated into NCCN databases and thereby made more accessible to NCCN staff.

Data Processing

Unlike the data from NPS sources, much of the data collected from external sources must undergo some degree of processing to meet the standards of the NCCN I&M Program; however some of the basic processing steps are very similar.

- All GIS data obtained from other entities are stored in the proper format, have the correct spatial reference information and have FGDC compliant metadata. This is especially applicable to data collected in support of the vital signs relating to land use/land cover change.
- All biodiversity data received from other entities should be entered into NPSpecies. This would include data sets like the Breeding Bird Survey. Also, if the data were taken from a report or published document, the reference must be entered into NatureBib.

Certain data sets will require more than the basic processing steps described above. The level of data procession required for external data sets such as those used in the vital signs monitoring program depends on the desired output. If more intensive analyses are intended than are provided through these means – such as conducting in depth analyses at a specific location rather than monitoring regional trends – the data may require additional processing steps. In such cases, the specific protocols should provide the necessary data processing requirements.

Remote sensing data sets such as satellite imagery or aerial photography will require varying levels of processing depending on how they are received. These may include geospatial processing or spectral processing. Ideally, all spatial data sets will be received in a geo-referenced format and may only require geographic transformations to meet NCCN standards.

Varying degrees of spatial and spectral processing may be necessary to adequately answer the proposed questions. The individual monitoring protocols outline the necessary processing steps.

Credits

This chapter was adapted from material developed by Geoffrey Sanders (National Capital Region).

Chapter 6 – Quality Assurance

We must have confidence in the data we use. Our analyses to detect trends and patterns in ecosystem processes require data of documented quality to minimize error and bias. Data of inconsistent or poor quality can result in loss of sensitivity and incorrect interpretations and conclusions. The potential for problems with data quality increases dramatically with the size and complexity of the data set (Chapal & Edwards 1994).

Our network data management system must ensure that our projects produce and maintain data of the highest possible quality. NCCN will establish and document protocols for the identification and reduction of error at all stages in the data lifecycle. These stages include project planning, data collection, data entry, verification and validation, processing, and archiving. Although a data set containing no errors would be ideal, the cost of attaining 95%-100% accuracy may outweigh the benefit. Therefore, we consider at least two factors when setting data quality expectations:

- frequency of incorrect data fields or records
- significance of error within a data field

We are more likely to detect an error when we work with clearly documented data sets and understand what a "significant" error is within a given data set. The significance of an error can vary with data sets and depends on where it occurs. For example, a two-digit number off by one decimal place is a significant error. A six-digit number, with the sixth digit off by one decimal place, is not a significant error. But one incorrect digit in a six-digit species number could indicate a different species. That is a significant error.

6.1 – NPS Mandate for Quality

Not long ago, maintaining data meant filling filing cabinets full of notebooks and paper. Now we are more likely to use computer hardware and software – technology that changes rapidly and sometimes unpredictably. If we expect our current data to be useful to future generations, the data must survive changes in technology. We can promote data longevity through high-quality documentation and maintenance during all phases of data management: during data collection, entry, verification, and validation. Well-documented data sets are especially important when sharing data.

NPS Director's Order #11B: "Ensuring Quality of Information Disseminated by the National Park Service," issued in 2002, promotes information and data quality. It defines *'quality'* as incorporating three key components–*objectivity*, *utility*, and *integrity*.

- *Objectivity* consists of: 1) *presentation*, which focuses on whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context, and 2) *substance*, which focuses on ensuring accurate, usable, and reliable information.

- *Utility* refers to the usefulness of the information to its intended users, from the perspectives of both the Network and the general public.

- *Integrity* refers to the security of information, e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Order #11B also specifies that information must be based on reliable data sources, which are accurate, timely, and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

High quality data and information are mandated by directives and orders, and they are vital to the credibility and success of the I&M program. Data need to meet national-level quality standards and need to be accessible to be used for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making.

6.2 – Quality Assurance and Quality Control Mechanisms

Palmer and Landis (2002) define *Quality Assurance* (QA) as "an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer." They define *Quality Control* (QC) as "a system of technical activities to measure the attributes and performance of a process, item, or service relative to defined standards." Quality Assurance procedures maintain quality throughout all stages of data development. Quality Control procedures monitor or evaluate the resulting data products.

QA/QC mechanisms are designed to prevent data contamination, which occurs when a process or event other than the one of interest affects the value of a variable and can introduce two fundamental types of errors into a data set:

- *Errors of commission* include those caused by data entry and transcription errors or malfunctioning equipment. They are common, fairly easy to identify, and can be effectively reduced upfront with appropriate QA mechanisms built into the data acquisition process, as well as QC procedures applied after the data have been acquired.
- *Errors of omission* often include insufficient documentation of legitimate data values, which could affect the interpretation of those values. These errors may be harder to detect and correct, but many of these errors should be revealed by rigorous QC procedures.

QA/QC procedures applied to ecological data include four procedural areas (or activities), ranging from simple to sophisticated, and inexpensive to costly:

1) defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata
2) checking for unusual or unreasonable patterns in data
3) checking for comparability of values between data sets
4) assessing overall data quality

Much QA/QC work involves the first activity (defining and enforcing standards), which begins with data design and continues through acquisition, entry, metadata development, and archiving. The progression from raw data to certified data implies increasing confidence in the quality of the data through time.

6.3 – Roles and Responsibilities

Quality assurance methods should be in place at the inception of any project and continue through all project stages to final archiving of the data set. It is critical that each member of the data management group work to ensure data quality. Everyone plays a part in producing and maintaining high quality data. Anyone assigned to a project is responsible for the quality of the results generated from his or her task(s).

The data manager is responsible for:

- developing protocols and SOPs to ensure data quality
- making project leaders, technicians, etc., aware of the established procedures and enforcing adherence to them
- evaluating the quality of all data and information against NPS standards before dissemination outside the network
- performing periodic data audits and quality control checks to monitor and improve the data quality program

Project leaders must:

- be aware of quality protocols and convey their importance to technicians and field crews
- ensure compliance with the protocols
- validate data after the verification process is complete
- review all final reports and information products

Technicians must follow established protocols for data collection, entry, and verification established in the inventory and monitoring protocol data management SOPs.

6.4 – Goals and Objectives

We must ensure that a project produces data of the right type, quality, and quantity to meet project objectives and user needs. Quality criteria should be set at a level proportionate to project-specific objectives. The EPA (2003) defines data quality objectives as qualitative and quantitative statements that:

- clarify the intended use of the data
- define the type of data needed to support management decisions
- identify the conditions under which the data are to be collected
- specify tolerable limits on the probability of making a decision error due to uncertainty in the data

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist the researcher in accurate data collection, entry, and validation. We will initiate a comprehensive set of SOPs and data-collecting protocols for quality control, field methodologies, field forms, and data entry applications with some built-in validation mechanisms.

Although specific QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the protocols for each monitoring vital sign, some general concepts apply to all network projects. The general QA/QC procedures presented in this plan were primarily adapted from the Draft Data Management Protocol (Tessler & Gregson 1997) and the ideas contained in Michener and Brunt (2000). These general guidelines will ensure that all data collected are checked for integrity before being integrated into the monitoring program databases. Refer to SOPs and monitoring protocols for specific QA/QC procedures.

6.5 – Data Collection

Careful, accurate recording of field observations in the data collection phase of a project will help reduce the incidence of invalid data in the resulting data set. Unlike a typographical error that occurs when a recorded observation is incorrectly transferred from a paper field form to a digital database, an incorrect entry in the field cannot be easily corrected. Therefore, attention to detail during data collection is crucial to overall data quality.

Paper field notebooks or data forms have been the primary methods for ecological data collection for many years. Although paper has advantages in terms of longevity and ease of use, it does not work well under some environmental conditions, and processing options are limited until the data are transferred to digital format. As an alternative to paper, several options for electronic data collection in the field are now available, including handheld computers, automated data collection instruments, and tape recorders.

Before the data collection phase of a project begins, the data manager is responsible for providing the protocols/SOPs for data collection and storage to the project leader. All field sheets and field data recording procedures must be reviewed and approved by the data manager and documented in the protocol SOPs. The project leader, in turn, will ensure that field crews understand the procedures and closely follow them in the field. If training is needed, the data manager will work with the project leader to provide that training. Field technicians are responsible for proofing raw data forms prior to data entry, ensuring their readability and legibility, and verifying and explaining any unusual entries. They are expected to understand the data collection forms, know how to take measurements, and follow the protocols.

6.5.1 – Methods for Reducing Collection Errors

Use a formatted, project-specific data sheet as opposed to a field notebook. When electronic data collection devices are not used, data should be recorded on paper data forms. We strongly recommend acid-free paper to prevent fading and subsequent data loss. Some circumstances may warrant the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions.

Standardized data sheets that identify the necessary information to be recorded and forms that reflect the design of the computer data entry interface will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant metadata such as date, collectors, weather conditions, etc.. They should clearly specify all required information, using examples where needed to ensure that the proper data are recorded. Data recorders should adhere to the following guidelines:

- All information added to the data sheet must be printed and clearly legible.
- If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased, and old information should not be overwritten.
- After error-checks of field forms are complete, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). The copies of the data sheets will be stored as specified in the protocol SOP, and the original data sheets will be used for data entry.

Use a handheld computer for data collection whenever possible. The use of handheld computers minimizes the need for manual data entry from field forms and associated transcription and data entry errors. Specially designed database or computer programs may be required for handheld computers, and the user interface should be customized to the project requirements. A customized data entry application has the advantage of incorporating on-the-spot QA/QC checks, so this data collection method probably provides the highest quality data.

These portable units, however, are subject to environmental hazards such as heat, dust, and moisture. When handheld computers are used for data entry in the field, the data should be downloaded daily to avoid potential loss of information. Thus, if a handheld unit fails during data collection, only the current day's data are lost. Batteries should be checked prior to a data collection trip, and they should be charged at the end of every field day. The use of a memory card that will store the data in case of damage to the unit or battery failure can prevent accidental loss of data. Also, in case the unit becomes inoperable in the field, printed data sheets should always accompany field teams on data collection trips.

Use automated data loggers where appropriate. Instruments with their own data acquisition systems are useful for collecting some types of data, such as water and air quality data. These devices can be calibrated and programmed to automatically record data and store them for later download directly to a computer, thereby eliminating the possibility for manual data entry errors.

Data loggers are an efficient means of recording continuous sensor data, but routine inspections are necessary. Moreover, environmental conditions, as well as power (e.g., sufficient battery charge) and maintenance requirements, are potential pitfalls when using these instruments. Regular downloads should be required since physical memory is usually limited.

Use a handheld tape recorder. Another alternative to paper field data forms is a handheld micro cassette tape recorder. Recorded observations are subsequently transcribed to paper or directly entered into computer files. As with other technological solutions, there are drawbacks including battery and tape maintenance, low environmental tolerance, and risk of failure. However, if a single data collector is in the field, tape recorders can provide an easily operated, efficient method of collecting high quality data.

Consider calibration, maintenance, and minimum timing requirements of field equipment. Accurate field measurements are possible only if field equipment is regularly calibrated and maintained. Where appropriate, consult reference manuals for recommended calibration and maintenance procedures. Once in the field, allow sufficient time for field equipment to adjust to its environment so it will record accurate measurements (for example, when using water quality probes and GPS units). Researchers should consider maintaining records of equipment calibration and failures that accompany their field data whenever possible.

Be organized and keep a log. Organization is the key to good data collection methods. Maintaining a log of important decisions and events will help clarify information and contribute to an accurate report.

Ensure that field crews receive proper training. Although protocols and SOPs are in place, they cannot guarantee that high quality data will be collected. Prior to routine data collection for a project, conduct training sessions to ensure that field personnel have a clear understanding of data collection procedures described in the SOPs. A training program may also include a process to certify that field staff understand and can perform the specified data collection procedures. The development of a training manual may be helpful for long-term monitoring data collection efforts and those that will involve a large number of field staff. Palmer and Landis (2002) provide an outline for a training manual and suggestions for planning training sessions.

Perform quantitative assessments of data quality. Repeating measurements is the primary tool for performing quantitative assessments of data. Project leaders should periodically review the work of field technicians to ensure that their work does not drift from standards during the course of the field season. Quantitative assessments may be considered if staff and funding are available, and Palmer and Landis (2002) describe several approaches that can be used.

6.6 – Data Entry

'Data entry' is the initial set of operations where we transfer raw data from paper field forms into a computerized form linked to database tables. Spreadsheets should not be used for data entry (data can be exported to a spreadsheet for manipulations post entry). When data are gathered or stored digitally in the field (e.g., on a data logger), data entry consists of the transfer of data (downloading) to a file in an office computer where they can be further manipulated.

Transferring data from field projects into the computer seems a fairly simple task. But the value of the data depends upon their accuracy, and we must feel confident about the overall quality of the data. Without proper preparation and some established guidelines, the quality and integrity of the data can be questionable. Ideally, data entry occurs as soon as possible–immediately after data collection is completed or as an on-going process during long projects–by a person who is familiar with the data. The primary goal of data entry is *to transcribe the data from paper records into the computer with 100% accuracy*. Yet, we know that a few transcription errors are unavoidable during data entry. Thus, all data should be checked and corrected during a data verification process.

The data manager, along with the project leader, should provide training in the use of the database to all potential users – especially those responsible for data entry and quality assurance. The project leader makes certain that data entry technicians understand how to enter data and follow the protocols. Data entry technicians are responsible for becoming familiar with the field data forms and differences in handwriting. They must also become familiar with the database software, database structure, and any standard codes for data entry used by the Network. At minimum, data entry technicians should know how to open a data entry form, create a new record, and exit the database properly. They must learn how to commit both a 'field' entry and a "complete record" entry and to correct mistakes made while typing.

6.6.1 – Methods for Reducing Data Entry Errors

Enter or download data in a timely manner. All data should be entered or downloaded into the project database as soon as possible, preferably at least once a week. Try to avoid delaying data entry until all the project data have been collected. Downloaded data should be periodically stored on CD or some other semi-permanent media.

Design efficient data entry forms and methods. A full-screen data entry form that mimics the field data forms can effectively reduce manual data entry errors due to the 1:1 correspondence of the attributes. A strategy to distinguish between validated data and newly entered data should be adopted. Data can be entered into an empty, fresh database table to avoid contaminating existing data and the new data appended to the master data only after formal verification, validation, and documentation. Alternatively, we can include validation attributes that indicate which data have been checked and validated by the project leader in the database. Regardless of strategy, we must clearly document the process for validation in the protocol data management SOP.

Build automated error checking features into the database. The most robust QA/QC measures for data entry should be built into the database design to perform automatic validation checks of data. Data entry forms reduce transcription errors through auto-filled fields, range limits, pick lists, and spelling checks. They provide controlled access to the database (i.e., forms are set for data entry only, which prevents accidental deletion or alteration of existing data). They control the sequence of data entry (i.e., certain fields require an entry before more information can be entered). They warn the operator when errors are made and provide an opportunity for correction before the data are committed to a file.

- *Auto-filled fields.* Whenever possible, the data in a field should be auto-filled by the computer. For example, if a location ID is composed of a park code, project code, and a unique number, those elements are automatically inserted into the location ID field, ensuring that the record always contains a unique identifier.
- *Range limits.* Where the appropriate values for a particular field span a finite range, the data entry program can check the entered value against the specified minimum and maximum values for that parameter. When a value is outside the accepted range, a warning message appears and asks the user to reenter a valid value. For some fields, values outside a specified 'normal' range may be acceptable. In this case, the warning message asks the user to verify the entry before continuing.
- *Pick lists.* The data entry application may also use pop-up pick lists for standardized text items where spelling errors can occur. For example, rather than typing in a species code or name (where a misspelling generates a new species in the database), the code or name is selected from a list of valid species codes or scientific names and automatically entered into the species field. A pick list may also be used when only certain entries are acceptable. Lists are not appropriate for all written fields but should be used when appropriate.
- *Unique constraints.* Duplicate and incorrect data entry can often be caught with the application of unique constraints on data entry fields. These constraints are particularly useful when importing data from other applications.

Provide a clean, organized work environment. Desktop space near the computer should be free of clutter and distractions that could cause the technician to lose her place. There should be enough space for two stacks of paper documents, one from which data are being entered and one from which data have been entered. A pad or notebook and some fine colored markers should also be available for making notes. (The need for a clean workspace also applies to the verification and validation phases.)

If possible, use two data entry technicians for data entry. When one technician reads the data from the field data forms and another enters them into the computer, the work is often faster and results in a lower error rate. If only one person is available, he should work at a slower pace to avoid errors. Like many monotonous tasks, data entry can be done in a personal rhythm that reduces the tediousness of the work.

6.7 – Verification and Validation Procedures

We appraise data quality by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices. *Data verification* checks that the digitized data match the source data, while *data validation* checks that the data make sense. It is essential that we validate all data as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a data set suspect. Although data entry and data verification can be handled by personnel who are less familiar with the data, validation requires in-depth knowledge about the data.

The data manager establishes SOPs for verification and validation and provides them to the project leader. Technicians will follow the SOPs for verification of data, make required changes, and document those changes. The project leader will ensure that the SOPs are followed. The project leader or designee will validate the data after verification is complete. The project leader is also responsible for reviewing all data products and reports before they are released outside the Network. The data manager and project leader will evaluate the results of verification and validation and determine any procedural or data form revisions that may be indicated by the results.

6.7.1 – Methods for Data Verification

Data verification immediately follows data entry and involves checking the accuracy of the computerized records against the original source, usually hard copy field records, and identifying and correcting any errors. When we have verified the computerized data as accurately reflecting the original field data, we can archive the paper forms and manipulate and analyze most data on the computer.

Among the following methods there is a direct correlation between effectiveness and effort. The methods that eliminate the most errors can be very time consuming while the simplest and cheapest methods will not be as efficient at detecting errors.

1) *Visual review at data entry.* The data entry technician verifies each record after input. She compares the values recorded in the database with the original values from the hard copy and immediately corrects any errors. This method is the least complicated since it requires no additional personnel or software. Its reliability depends entirely upon the person keying data and thus, is probably the least reliable data verification method.
2) *Visual review after data entry.* All records are printed upon the completion of data entry. The values on the printout are compared with the original values from the hard copy. Errors are marked and corrected in a timely manner. This method's reliability increases if someone other than the person keying data performs the review. Alternatively, two technicians can perform this review. One technician reads the original data sheets (the reader), and the second reads the same data on the printout (the checker).
3) *Duplicate data entry.* The data entry technician completes all data entry, as normal. Random records are then selected (every **n**th record) and entered into an empty replica of the permanent database, preferably by someone other than the person keying the permanent data. Then a query is used to automatically compare the duplicate records from the two data sets and report any data mismatches. The data technician manually reviews any disparities and makes corrections if necessary. This method adds the overhead of retyping the selected records, as well as the creation of a comparison query, but it becomes increasingly successful as the value of **n** decreases. Professional data entry services frequently use this method.

We can also use the entered data to calculate simple summary statistics with statistical software. These summary statistics can help us catch a duplicate or omitted entry. For example, we can view the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample. We can pose the same question in different ways; differences in the

answer provide clues to errors. The more checks we devise to test the completeness of the data, the greater our confidence that we have completely verified the data.

To minimize transcription errors, our policy is to verify 100% of records to their original source by permanent staff. In addition, 10% of records are reviewed a second time by the project leader, and we report the results of that comparison with the data. If the project leader finds errors in her review, then we verify the entire data set again.

Although we may have correctly transcribed the data from original field notes or forms, they still might be inaccurate or illogical. For example, entries of stream pH of 25.0 or a temperature of 95°C in data files raise doubt about their accuracy; and such entries almost certainly are incorrect, whether or not they were properly transcribed from field forms. This process of reviewing computerized data for range and logic errors we call *validation*, and it can accompany data verification *only* if the operator has comprehensive knowledge about the data. More often, validation is a separate operation carried out *after* verification by a project specialist who can identify generic and specific errors in particular data types. Corrections or deletions of logical or range errors in a data set require notations in the original paper field records about how and why the data were changed. Modifications of the field data should be clear and concise while preserving the original data entries or notes (i.e., no erasing!). Validation efforts should also include a check for the completeness of a data set since field sheets or other sources of data could easily be overlooked.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Nevertheless, validation is a critically important step in the certification of the data. Invalid data commonly consist of slightly misspelled species names or site codes, wrong dates, or out-of-range errors in parameters with well defined limits (e.g., elevation). But more interesting and often puzzling errors can be detected as unreasonable metrics (e.g., stream temperature of 70°C) or impossible associations (e.g., a tree 2 feet in diameter and only 3 feet high). We call these types of erroneous data *logic errors* because using them produces illogical (and incorrect) results. The discovery and correction of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Histograms, line plots, and basic statistics can reveal possible logic and range errors.

6.7.2 – Methods for Data Validation

The following general methods can be used to validate data. Specific procedures for data validation depend upon the vital sign being monitored and will be included in the monitoring protocols.

Data entry application programming. Certain components of data validation are built into data entry forms. The simplest validation during data entry is range checking, such as ensuring that a user attempting to enter a pH of 20.0 gets a warning and the opportunity to enter a correct value between 1.0 and 14.0 (or better yet, within a narrow range appropriate to the study area). Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage are important.

Edwards (2000) suggests the use of 'illegal data' filters, which check a specified list of variable value constraints on the master data set (or on an update to be added to the master) and create an output data set. This output data set includes an entry for each violation, along with identifying information and an explanation of the violation. He illustrates the structure of such a program, written in the SAS® programming language.

A caveat should be interjected regarding the operative word 'illegal'. Even though a value above or below a given threshold has never before been observed and the possibility that it could occur seems impossible, such an observation is not always an illegal data point. Edwards (2000) points out that one of the most famous data QA/QC blunders to date occurred when NASA's computer programs deleted satellite observations of ozone concentrations that were below a specified level, seriously delaying the discovery of the ozone hole over the South Pole.

Outlier Detection. According to Edwards (2000), "the term outlier is not (and should not be) formally defined. An outlier is simply an unusually extreme value for a variable, given the statistical model in use." Any data set will undoubtedly contain some extreme values, so the meaning of 'unusually extreme' is subjective. The challenge in detecting outliers is in deciding how unusual a value must be before it can (with confidence) be considered extreme.

Data quality assurance procedures should not try to **eliminate** outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is a better way to explain this quality assurance goal. If contamination is not detected during data collection, it is usually only detected later if an outlying data value results. When we detect an outlier, we should try to determine if some contamination is responsible.

We can use database, graphic, and statistical tools for ad-hoc queries and displays of the data to detect outliers. Some of these outlying values may appear unusual but prove to be quite valid after confirmation. Noting correct but unusual values in documentation of the data set saves other users from checking the same unusual values.

Other exploratory data analyses. Palmer and Landis (2002) suggest that in some cases, calculations for assessments of precision, bias, representativeness, completeness, and comparability may be applicable and that for certain types of measurements, evaluation of a detection limit may also be warranted (the authors provide examples of procedures that may be applicable). Normal probability plots, Grubb's test, and simple and multiple linear regression techniques may also be used (Edwards, 2000; the author provides SAS and S-Plus code for constructing normal probability plots and examples of output showing normal and non-normal distributions).

6.8 – Version Control

Our network manages files from a multitude of sources, comprising many formats with many iterations of a particular product. Some of the files are complete, some are works-in-progress, and for others the status cannot be determined. In addition to files it collects, NCCN also

generates many files, some of which fall into the complete, works-in-progress, and undetermined status categories. Determining the status of a single file can be difficult, but determining the current file within a series of similarly named files can be almost impossible.

Version control is the process of documenting the temporal integrity of files as they are being changed or updated. Change includes any alteration in the structure or content of the files, and such changes should not be made without the ability to undo mistakes caused by incorrect manipulation of the data. Whenever we complete a set of data changes, we should save the file with a unique name, a simple act that should become routine for all data handlers.

Prior to any major changes to a file, we should store a copy of the file with the appropriate version number. This allows the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used in any analysis.

The data manager determines the version control method that will be used, and other network personnel are responsible for accurately designating versions for any files upon which they have worked. Software tools that assist in file management can be helpful–for example, to creating databases that include fields to record the file's revision history. Backup routines can be built into the databases that allow for automatic file renaming and archiving. Important program files can be catalogued in a simple index or more formally tracked and archived using professionally developed version control software.

6.8.1 – Version Control Options

Dates. Using a date provides logical version control. The date is usually formatted as YYYYMMMDD, where DD is optional depending on the frequency of changes. Indicating the month as a three-character text string (e.g., 2005Sep12) increases readability and avoids confusion over which digits signify year, month and date.

Sequential numbers. We can designate versioning of archived data sets by adding a number to the file name, e.g., 001 or V1.0 for the first version. We assign each additional version a sequentially higher number. We should also document the date that a file becomes a new version, perhaps through assignment of database folder (or directory) names.

Version control software. We can eliminate the work of differentiating multiple versions of documents by using version control software to append modifying characters to the file name. Such software applications track changes made to a document, add comments related to the different document iterations, and retrieve the document at any recorded stage of development. These applications are available in either desktop or online formats.

Version control software should:

- Preserve previous versions of documents for possible recovery
- Track documents as they change during the course of the developmental and editorial phases of document/report creation

- Prevent conflicts between multiple collaborators by prohibiting multiple edits to the same file at the same time
- Evaluate the document creation process by tracking who changes a file, when they make the change, and what changes they make
- Reduce storage requirements by eliminating multiple copies of complete documents

6.9 – Data Quality Review and Communication

The National Park Service requires QA/QC review prior to communicating/disseminating data and information. Only data and information that adhere to NPS quality standards can be released.

Director's Order #11B states that all information (e.g., brochures, research and statistical reports, policy and regulatory information, and general reference information) distributed by the NPS (including information obtained from sources outside of the NPS) must be accurate, reliable and timely in nature. Therefore, we must evaluate and identify the types of information that will be subject to the guidelines. Information disseminated to the public must be approved by the appropriate reviewing officials and programs. The QA/QC standards used in producing the information and that substantiate the quality of the information must be formally documented. Also, mechanisms must be in place for receiving and addressing comments/complaints pertaining to the quality of data.

Data are distributed to the public through the NCCN I&M web page, national web sites such as the Biodiversity Data Store and the Natural Resource/GIS Data Store, and public access databases such as NPSpecies and NatureBib. Any information distributed through any of these mechanisms must undergo internal QA/QC procedures and be approved for release.

6.9.1 – Data Quality Review Methods

The NCCN Inventory and Monitoring Program will establish guidelines and protocols to ensure compliance with DO #11B. These protocols will document both internal and external review procedures for data and information disseminated outside the network, as well as a system for processing complaints about data quality.

Edwards (2000) suggests the initiation of quality circles, regular meetings of project leaders, the data manager, and data management personnel for discussing data quality problems and issues. These meetings promote teamwork attitudes while focusing brainpower on data quality issues. Participants become more aware of quality issues and learn to anticipate problems. Moreover, all participants develop a greater appreciation of the importance of their role in data quality and the entire monitoring effort.

6.9.2 – Value of Feedback from QA/QC Procedures

Quality assurance procedures may need revision to improve data quality if random checks reveal unacceptable levels. Quality checks should not be performed with the sole objective of eliminating errors, as the results may also prove useful in improving the overall quality assurance process. For example, if the month and day are repeatedly reversed in a date field, the data entry

technicians may require retraining about the month/day entry order. If retraining is unsuccessful in reducing the error's occurrence, the computer program may need to be rewritten so that month and day are entered separately, field length limits are enforced, or a pick list is created. In this manner, the validation process will serve as a means of improving quality as well as controlling the lack of quality.

We can modify field data forms to avoid common mistakes when necessary. With knowledge of validation errors and exploratory data results in hand, the field data forms as the source of the logic errors can be more easily re-evaluated. Often minor changes, small annotations, or adding check boxes to a field form can remove ambiguity about what to enter on the form. Perhaps surprisingly, when we find the same type of validation error occurring repeatedly in different data sets, the field form–not the field crew–is usually at fault. Repeated errors found during validation can also mean that protocols or field training are at fault, which can then be corrected.

6.9.3 – Monitoring Conformance to Plans and Standards

Data managers can use periodic data audits and quality control inspections to maintain and improve their data quality program. They must verify that staff is operating in conformance with the data quality procedures specified in this plan and the protocol specific data management plans. They should track and facilitate the correction of any deficiencies. These quality checks promote a cyclic process of continuous feedback and improvement of the both the data and quality assurance process.

Periodic checks by the data manager to see if network staff are adhering to the data quality procedures established in the Data Management Plan and protocol SOPs may include verification of the following:

- Data collection and reporting requirements are being met
- Data collection and reporting procedures are being followed
- Verification and validation procedures are being followed
- Data file structures and maintenance are clear, accurate and according to plan
- Revision control of program documents and field sheets are adequate
- Calibration and maintenance procedures are being followed
- Seasonal and temporary staff have been trained in data management practice
- Metadata collection and construction for the program proceeds in a timely manner
- Data are being archived and catalogued appropriately for long term storage

The results of quality assessments should be documented and reported to the research staff and the network coordinator. The project leader and coordinator are responsible for ensuring that non-conformities in data management practices are corrected.

6.9.4 – Communicating Data Quality

The Network will use data documentation and metadata to notify end users, project leaders, and network management of data quality. A descriptive document for each data set/database will provide information on the specific QA/QC procedures applied and the results of the review.

Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

Credits

This chapter was adapted from material developed by Deborah Angell (Sonoran Desert Network). Special thanks to Doug Wilder (Central Alaska Network) and Sara Stevens (Northeast Coastal and Barrier Network) for contributions and ideas.

Chapter 7 – Data Set Documentation

Note: This chapter is currently in revision as of September 2005.

7.1 – Purpose of Metadata

Data documentation is a critical step toward ensuring that data sets are useable for their intended purposes well into the future. This involves the development of metadata, which can be defined as information about the content, quality, condition and other characteristics of data. Additionally, metadata provide the means to catalog data sets within intranet and internet systems, thus making the data available to a broad range of potential users.

Data sets seem to sometimes take on lives of their own. Some seem to have the ability to reproduce and evolve on multiple hard drives, servers and other storage media. Others are masters at remaining hidden in digital formats or in forgotten file drawers. In addition, once data are discovered, a potential data user is often left with little or no information regarding the quality, completeness, or manipulations performed on a particular "copy" of a data set. Such ambiguity results in lost productivity as the user must invest time tracking information down or, in the worst case, the data set is rendered useless because answers to these and other critical questions cannot be found. As such, data documentation must include an upfront investment in planning and organization.

While the importance of metadata is universally accepted within the data management community, the approaches for collection and levels of detail are varied (sometimes referred to as the "101 ways"). However, following are some important considerations for the development of data documentation strategies.

- Executive Order 12906, signed by President Clinton in 1994, mandates federal agencies to "...document all new geospatial data it collects or produces, either directly or indirectly..." using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). In addition, EO 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.
- The FGDC Biological Data Profile contains all the elements of the CSDGM and includes additional elements for describing biological data sets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data sets is recommended.
- All GIS data layers must be documented with applicable FGDC and NPS metadata standards (NPS Metadata Profile).
- While there are numerous tools available for developing metadata, the NPS Integrated Metadata System Plan is limited to three recommended desktop applications: Dataset Catalog, ArcCatalog, and Spatial Metadata Management System (SMMS).

7.2 – NPS Integrated Metadata System Plan and Tools

As noted above, the NPS Integrated Metadata System Plan is limited to three recommended desktop applications for collecting metadata. These include Dataset Catalog (developed by the

I&M Program), and two commercial off the shelf metadata tools, ArcCatalog and SMMS. The following is a brief description of each of these tools, including their potential utility in metadata creation. A fourth tool, the Metadata Parser (mp, developed by USGS) and its utility in metadata creation is also briefly discussed.

Dataset Catalog: Dataset Catalog is a tool for cataloging abbreviated metadata on geospatial and biological data sets pertaining to park(s) and/or a network. It provides parks and/or networks a means whereby they can inventory, organize, and maintain information about data set holdings locally. While Dataset Catalog is not intended to be an exhaustive metadata listing, it does assist parks and networks in beginning to meet the mandates of EO 12906. With the current version of Dataset Catalog (version 2), records can be exported as an FGDC text file, which can than be imported into other metadata tools. Version 2.1 (in development) will include the ability to export records in Extensible Markup Language (XML). The I&M Program recommends that all relevant data sets at I&M parks and networks be cataloged in at least simple Dataset Catalog format. Our plan is plan to follow this recommendation.

Spatial Metadata Management System: SMMS is a tool with the capability to create, edit, view, and publish metadata that is compliant with FGDC requirements. SMMS uses an MS Access database structure combined with an advanced FGDC-compliant metadata editor. The software allows selection of views depending on whether the user wants the full standard, biological, or the minimal compliant view of Sections 1 and 7. There is online Help to describe the purpose, usage or mandatory status of metadata elements. The context-sensitive help file provides the FGDC definition for each field on the screen. In addition to Help files, there are sample metadata records for most sections that provide "real world" examples. The NPS Integrated Metadata System Plan recommends SMMS for FGDC Biological Profile and other geospatial metadata creation. We are currently using SMMS, although this platform may be dropped in the near future as customization of ArcCatalog metadata tools makes it the optimal tool.

ArcCatalog: ArcCatalog is a management tool for GIS files contained within the ArcGIS Desktop suite of applications. With ArcCatalog, users can browse, manage, create, and organize tabular and GIS data. In addition, ArcCatalog comes with support for several popular metadata standards that allow one to create, edit, and view information about the data. There are editors to enter metadata, a storage schema, and property sheets to view the data. With ArcCatalog users can view GIS data holdings, preview geographic information, view and edit metadata, work with tables, and define the schema structure for GIS data layers. Metadata within ArcCatalog is stored exclusively as Extensible Markup Language (XML) files. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata. An optional but highly recommended extension for ArcCatalog is the NPS Metadata ArcCatalog Extension developed by NPS Midwest Region GIS Technical Support Center. The extension fixes several ArcGIS 8 metadata errors and provides added functionality for NPS users. Development is also underway to provide Biological Profile editing capability and NPS Profile support.

Metadata Parser: The MetaParser (mp) program is used to validate metadata records by checking the syntax against the CSDGM and to generate compliant output files for posting to

clearinghouses. It generates a textual report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard.

Until recently many NPS data stewards collected, parsed and stored metadata (and GIS data sets) in the NPS GIS Clearinghouse managed by North Carolina State University (NCSU). However, efforts are currently underway to unify and streamline metadata development. This new approach utilizes existing desktop metadata creation applications, as well as an online integrated metadata database (NR-GIS Metadata) and a web based data server (NR-GIS Data Server). NR-GIS Metadata and Data Store will comprise a web based system to integrate both data dissemination and metadata maintenance. It will be possible to update Dataset Catalog records in the NR-GIS Metadata database or in the source desktop application (i.e., ArcCatalog, Dataset Catalog, SMMS). Non-sensitive NR-GIS Metadata records are automatically posted to NPS Focus.

7.3 – Metadata Process and Workflow

Note: SOPs and guidance documents are under development for utilizing NPS Metadata Editor and ArcCatalog.

7.3.1 – Identify Relevant Data Sets and Compile Pertinent Metadata

NCCN data can be grouped, at least initially, in three broad categories based on origin. These categories include legacy data sets (primarily data collected prior to the inception of the NCCN I&M program), non-programmatic data sets (ongoing data collection efforts conducted by entities outside the scope of the I&M program), and I&M data sets.

Legacy Data

In many cases, legacy data are initially identified as part of data mining efforts. Unfortunately, many of the legacy data sets will be missing pertinent information, and the originator may no longer be in contact. Thus, an "adequate" level of documentation may not be possible. However, the data and all supporting documentation related to it should be assembled and reviewed (see chapter 5 for further information on legacy data acquisition and processing). Many legacy data sets will need to be converted to a standard database format for incorporation and future analyses. Data entry, validation and verification procedures will follow those contained within this DMP. A processing and revision log will be maintained with the data set for capture of pertinent metadata.

Non-programmatic Data

The I&M program is not the only entity gathering relevant inventory and/or monitoring data pertinent to park management. We will make every effort to capture and assimilate all relevant data. The outside entity will be contacted, and a request will be made for available metadata and/or a metadata interview will be conducted. As with legacy data, data files may need to be converted to a standard database format for analysis. Data entry, validation and verification procedures will follow those contained within this DMP. A processing and revision log will be maintained with the data set for capture of additional metadata.

I&M Data
For new projects, metadata development will begin up front, with an interview with the principal investigator to explain what will be needed to properly document the data. In most instances, this will include completion of a basic metadata survey for inclusion in the data manager's project file, as well as submission of supporting documentation (proposal, SOPs, etc.). In addition, a database structure will be developed by, or in close consultation with, the data manager, to ensure compliance with the principles and procedures contained within this DMP. Updates and revisions to the metadata will be conducted in tandem with data submissions.

7.3.2 – Create Dataset Catalog Record

Because metadata documentation can be accomplished in a variety of formats and levels of detail, it can become a consuming, some might say exhaustive task. As a starting point, we will develop a simple Dataset Catalog record for relevant spatial and non-spatial data. This approach provides brief metadata for all NCCN data holdings in a searchable, centralized location. In addition, managers can identify and prioritize data sets for which formal metadata will be developed and identify the status of metadata documentation for a particular data set (i.e., planned, in work, complete). These records can be imported into the online NR-GIS Metadata System or continue through additional processing steps based on data type, source, and importance.

Prioritization of data sets for further documentation will be based upon current or anticipated future use. In other words, data sets that will be used repeatedly in analysis or with high probability for data sharing will be addressed first. All GIS layers will be documented with applicable FGDC and NPS metadata standards.

7.3.3 – Select Metadata Tool and Complete Record

Historically, parks in the NCCN and Pacific West Region have used the Fire and Mountain Ecology Bioinformatics Lab at the University of Washington through CESU agreements for completion of FGDC compliant metadata and posted parsed records to the clearinghouse at the Metadata Clearinghouse on the Fire and Mountain Ecology website (http://www.cfr.washington.edu/research.FME/research/metadata.htm). However, with recent improvements to centralize metadata efforts and customized metadata tools, we plan to shift over to in house metadata creation using ArcCatalog as the primary metadata tool with extensions to incorporate biological profile information. The review of these new tools will be underway during FY05.

7.3.4 – Make Information Available

At a minimum, metadata and associated data will be submitted to NR-GIS Metadata and Data Store. This will be accomplished utilizing the recommended desktop applications. Additionally, information on data holdings should be conveyed in a meaningful manner for park resource managers, researchers, and others with a potential interest/stake in park management and/or research endeavors. Similar to metadata creation the mechanisms and formats for accomplishing this are varied. In addition to FGDC Text files, Dataset Catalog can output a list of all records, single record reports, and/or a data dictionary report. Depending on the target audience, these

standardized outputs can be useful in conveying information on program data holdings and summaries of database structures. Obviously, customized queries and reports can also be generated. Other standardized outputs include ArcCatalog stylesheets. The NPS Metadata ArcCatalog Extension contains custom stylesheets, which can be invoked from the metadata toolbar. These can be utilized to depict pertinent details in a more coherent format than standard metadata outputs.

Credits

This chapter was adapted from material developed by Teresa Leibfreid (Cumberland Piedmont Network) and Bill Moore (Mammoth Cave National Park). Special thanks to Doug Wilder (Central Alaska Network) for significant contributions and ideas.

Chapter 8 – Data Analysis and Reporting

Sound data management practices are the key to having a credible and useful monitoring program. However, a successful monitoring program requires that data are analyzed, interpreted, and provided to managers and others at regular intervals and in a format appropriate for each audience.

8.1 – Types of Reports

There are several types of documents that result from reporting and analysis:

- Annual reports
- Analysis and synthesis reports
- Scientific journal articles and book chapters
- Interpretation and outreach brochures and displays

It is important to indicate at project inception the types of reports that will be generated and who will be responsible for generating each. These considerations clarify expectations for these reports and ensure that there is sufficient program accountability, documentation, and evaluation.

The NCCN Monitoring Plan presents an overview for monitoring projects whose implementation will be funded in the near-to mid-term. The Implementation Plan also includes timetables and schedules for report generation related to each project, the types of analyses included, and peer review requirements. More detailed description of data analysis will also be contained in the protocols associated with each monitoring project.

8.1.1 – Annual Reports

The major purposes of annual reports are to:

- summarize annual data and document monitoring activities for the year
- describe current condition of the resource being monitored
- document changes in monitoring protocols
- increase communication within the park and network.

Many of our monitoring programs will be active each year, and those programs will generate annual reports each year. However, some sampling regimes do not require annual activity. For example, some protocols collect data every third year, every fourth year, or even every tenth year. Those programs will produce "annual" reports only when there are significant monitoring activities to document. These reports will be generated from automated data analyses developed for each monitoring project.

- *Audience*: Network staff, park staff including administration, scientists working in parks
- *Review*: internal network review

8.1.2 Analysis and Synthesis Reports

The role of analysis and synthesis reports is to:

- determine patterns/trends in condition of resources being monitored
- discover new characteristics of resources and correlations among resources being monitored
- analyze data to determine amount of change that can be detected by this type and level of sampling
- provide context, i.e. interpret data for the park within a multi-park, regional or national context
- recommend changes to resource management (feedback for adaptive management).

These reports can provide critical insights into resource status and trends, which can then be used to inform resource management efforts and regional resource analyses. This type of analysis, more in depth than that of the annual report, requires several seasons of sampling data. Therefore, these reports are not written more frequently than every three to five years for resources sampled annually. For resources sampled less frequently, or which have a particularly low rate of change, intervals between reports may be longer.

It is important that results from all monitoring projects within and across all NCCN parks be integrated across disciplines in order to interpret changes to park resources. This will be accomplished with a network synthesis report produced at no more than 10-year intervals.

- *Audience*: superintendents, park resource managers, network staff, and external scientists
- *Review*: external, blind peer-review with at least 3 subject-matter experts including one statistician

8.1.3 – Scientific Journal Articles and Book Chapters

This aspect of the program will be directed by the program managers, and is more at their discretion than previous reports. Publishing scientific journal articles and book chapters is primarily conducted to communicate advances in knowledge and is a very important, widely acknowledged means of quality assurance and quality control, via the academic peer-review process. Putting a program's methods, analyses, and conclusions under the scrutiny of a scientific journal's peer-review process is one of the best ways to ensure scientific rigor. This may be an important role for USGS.

Scientific journal articles and book chapters produced by NCCN efforts are tracked by the NCCN monitoring program; new publications are part of the Annual Administrative Report and Work Plan, which is sent to the regional and national offices each year. Additionally, all scientific journal articles and book chapters will be entered into the NatureBib database. Principle Investigators of recently published work in the NCCN frequently make presentations at professional workshops and conferences and will be invited to present their findings at Technical Committee and Board of Directors meetings.

- Audience: scientific community
- Review: peer-review conducted by journal or book editor

8.1.4 – Interpretation and Outreach

Scientific information gained from monitoring programs usually requires a concerted effort to be translated for the general public. Oral presentations include Science Days, which is a program led by the North Cascades Science Advisor who invites Network researchers to present their work and results and to share this information with the interested public. Network staff members also speak at trainings for seasonal employees and to special interest groups (e.g., Washington Native Plant Society, elder hostels, Olympic Park Institute, etc.). Network staff share discoveries with the public in written form by contributing articles to natural history newsletters and other media in the parks which produce them. In the future, the Network plans to produce brochures and fact sheets regarding monitoring.

Interpretation and outreach is a perfect place for the NCCN Vital Signs Monitoring program to team up with the NCCN Research Learning Center (RLC). The RLC promotes research in parks, as well as acting as a bridge between scientists and the public. The NCCN Network Coordinator and NCCN program leads are working with the RLC program to form connections with college students, partners, and the interested public to provide Vital Signs Monitoring program information to the community in a digestible form.

- Audience: the interested public
- Review: reviewed by project leads for accuracy

8.2 – Identifying Project Objectives and Data Products

The development of data products will be guided by project objectives and when available, by the protocols and data management and data analysis SOPs (*guidance documents in development*).

Project scopes-of-work, and whenever possible, annual reports from previous years will be used to identify the tables and fields necessary for data summary. Assuming a database already exists, one can begin by identifying the core or critical tables with the desired data addressing the project goals. Then one may review the total database design including Entity-Relationship Diagrams (ERDs), priority fields, and the tables that contain them, to work towards building queries for data summary and analysis. These are also the preliminary steps that will be used for report automation (below).

Individual fields will be identified which provide the source data for the desired analyses. Very often, the fields are to come from two or more tables. So it will become necessary to join tables using queries. ERDs will be used to determine the most effective way to write queries. Usually analysis queries do not yet exist, and these will be built by the Data Manager in coordination with the Project Leader. The Project Leader will start by identifying one or two variables to be used in the analysis, and then add independent variables that will be used for grouping purposes

such as site, habitat, or time interval. These will form the basis for the analysis queries. The analysis will be developed through a series of queries until either the data are exported to outside applications (see below) or the desired result is calculated.

8.3 – Automated Reporting and Data Summaries

The automation of data summaries and annual reports facilitates our network's ability to manage multiple projects. The Network uses Microsoft Access to automate its data summaries and reports although the process would be basically the same in other application environments.

8.3.1 – Annual Report as Template

The development of automated reporting would be greatly facilitated by existing annual reports when these are available. If not, a detailed annual report template will be worked out between the Project Leader and the Data Manager. The job of the Data Manager will be to take the extracted tables and fields and to process them to create output that is similar, if not identical, to the existing annual report template. This task requires that the Project Leader have a very clear idea of what will be needed for the annual reports. Usually, the Data Manager will need only to develop the necessary queries, macros, modules and reports. The process of automation is then to link the various database objects to a simple user-interface.

8.3.2 – Project Leader Interviews

Often, the Project Leader will need to modify the existing annual reports to better meet the requirements of Vital Signs projects. The Data Manager will plan on expanding and improving the analysis and reporting products during the automation phase, if possible, because the underlying database objects will have to be thoroughly reviewed anyway. Developing the automated reporting system beyond the existing annual reports will require frequent and repeated interviews by the Data Manager with the Project Leader. Once again, the project objectives will facilitate communication between the Data Manager and the Project Leader. The Network will often sketch out hardcopies of the data summaries, annual reports and user-interfaces in pencil prior to starting the implementation phase.

8.3.3 – Automated Report Generation

Following discussions on data summary and reporting requirements, the next step is to implement the automated report. Except in the simplest of situations, we will develop each analysis independently through a series of Access queries. The resultant query will be loaded as one of several controls (textboxes) into the summary report. All queries that feed into a single report may be triggered from a form button, a macro, or a combination of both. The Network will use modules and VBA for automation only to the extent that more complex automation applications are required.

8.4 – Exporting Data for Analysis and Exploration

At times, data will need to be exported out of the database to other software applications. The Network is planning to export data from Access databases for most statistical analysis beyond means, standard deviations, and other descriptive statistics. NCCN will use third party statistical software for frequency distribution plots, tests for normality and analysis of variance such as SAS, SPSS and NCSS. Other external software requiring data exports will most likely include special application software such as GS+ for geostatistical analysis and Distance 4.0 for estimating animal population abundance using distance sampling.

Analysis objectives will be reviewed by the Data Manager and Project Leader together prior to attempting data exports. The field order [order of variables in the resulting 'flat' file] is most easily controlled while the data are still in the relational database. Therefore, the Data Manager will assist in setting up the necessary queries to provide the field order. Field types are typically determined early on, when the database is being designed. To minimize risk of data loss as a result of data type conversion, changes in data types will be kept to a minimum once monitoring projects are underway.

8.5 – Making Data Available for Third-Party Analysis

In order to create a data flow for subsequent analysis by third parties, the Network will establish a timeline and data processing steps including error-checking, summarizing, analyzing, and distributing data. A data flow model including error-checking, analyzing, reporting, and distribution is described in detail in Chapter 4. What follows are the steps and timeline involved in having each project leader or the data manager complete these tasks, and make their particular data set available to others each year. (*In development*)

Credits

This chapter was adapted from concepts and material developed by Gareth Rowell and Mike Williams (Heartland Network and Prairie Cluster Prototype), and Dorothy Mortenson (Southwest Alaska Network).

Chapter 9 – Data Distribution and Dissemination

9.1 – Data Ownership

9.1.1 – National Park Service Policy on Data Ownership

The National Park Service defines conditions for the ownership and sharing of collections, data, and results based on research funded by the United States government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the NPS:

- All data and materials collected or generated using NPS personnel and funds become the property of the NPS.

- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.

- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

The Office of Management and Budget (OMB) ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. Specifically, OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36 of Circular A-110, "Intangible Property" describes the following administrative requirements pertinent to data and ownership:

> *(a) The recipient* (i.e., organization receiving federal monies for natural resource inventory and/or monitoring) *may copyright any work that is subject to copyright and was developed, or for which ownership was purchased, under an award. The Federal awarding agency(ies*) (i.e., the NPS) *reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.*
>
> Section 36 also states:
>
> *(c) The Federal Government has the right to:*
>
> > *(1) obtain, reproduce, publish or otherwise use the data first produced under an award*

(2) authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes

(d) (1) In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA (5 U.S.C. 552(a)(4)(A)).

(2) The following definitions apply for purposes of paragraph (d) of this section:

(i) Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples)…

(ii) Published is defined as either when:

(A) Research findings are published in a peer-reviewed scientific or technical journal; or

(B) A Federal agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

(iii) Used by the Federal Government in developing an agency action that has the force and effect of law is defined as when an agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

9.1.2 – Establishing Data Ownership Guidelines

To ensure that proper ownership, format, and development of Network products is maintained, all cooperative or interagency work should be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving NCCN should include OMB Circular A-110 cited under the *Reports and Deliverables* Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

> "As the performing organization of this agreement, *institution or organization name* shall follow the procedures and policies set forth in OMB Circular A-110."

Every cooperative or interagency agreement or contract should include a list of expected deliverables and products clearly defined within each agreement or contract. Details on formatting and media types that will be required for final submission should also be included. Expected products include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

The following statement must be included in the Reports and Deliverables section of all NCCN agreements and contracts:

"All reports and deliverables must follow the most recent version of the North Coast and Cascades Network Product Specifications."

Researchers should also provide a schedule of deliverables that includes sufficient time for NPS review of draft deliverables before scheduled final submissions.

9.2 – Data Distribution

One of the most important goals of the I&M Program is to integrate natural resource inventory and monitoring information into NPS planning, management, and decision making. To accomplish this goal, procedures should be developed to ensure that relevant natural resource data are made available for management decision-making, research, and education. Providing well-documented data in a timely manner to park managers is especially important to the success of this program. We will work to make certain that:

- Data are easily discoverable and obtainable
- Data that have not yet been subjected to full quality control will not be released unless required in response to a FOIA request
- Distributed data are accompanied by complete metadata that clearly establishes the data as a product of the NPS I&M Program (where appropriate)
- Sensitive data are identified and protected from unauthorized access and inappropriate use
- A complete record of data distribution/dissemination is maintained

To accomplish this, we will use a number of distribution methods that will allow information collected and developed as part of the program to become widely available to park staff and the public.

9.2.1 – Data Distribution Mechanisms

The primary mechanism for distributing data products will be the internet. This will allow the data and information to reach a broad community of users. In support of the NPS I&M Program, web-based applications and repositories have been developed to store a variety of park natural resource information. We will use the following applications and repositories to distribute our data products:

- **NatureBib** – a master web-based database housing natural resource bibliographic data for parks (NatureBib Home Page)

- **NPSpecies** – a master web-based database to store, manage and disseminate scientific information on the biodiversity of all organisms in all NPS units (NPSpecies Home Page)

- **Biodiversity Data Store** – a digital archive of documents, GIS and non-GIS data files that document the presence/absence, distribution and/or abundance of any taxa in NPS units (Biodiversity Service Center Home Page)

- **Natural Resource and GIS Metadata and Data Store -** online repository for metadata and associated data products. (NPS NR-GIS Metadata and Data Store Home Page)

- **North Coast and Cascades Network Website** – provides detailed information about the network and its I&M Program. Metadata on all inventory and monitoring products developed as part of the Network's I&M plan will be posted to this site. Data and products will either be available through the site or users will be directed to where the data are stored (North Coast and Cascades Network Home Page).

Table 9.1 below provides a list of types of information that will be uploaded to these sites.

Web Application Name	Data types available at site
NPSpecies	Data on park biodiversity (species information)
NatureBib	Bibliographic citations related to park resources
Biodiversity Data Store	The raw or manipulated data and products associated with Inventory and Monitoring data that have been entered into NPSpecies
NR-GIS Metadata and Data Store	Metadata and spatial/non-spatial data products
NCCN Website	Reports and metadata for all I&M projects administered by NCCN

Currently, the NR-GIS Metadata and Data Store and the Biodiversity Data Store are under development. Until procedures and further guidance become available for the use of these two repositories, we will disseminate all I&M data via the NCCN website. When both repositories are completely operational, the Network will upload all applicable data and information to each of those sites as needed.

Because network data will reside in the repositories listed above, this data will automatically be searchable via the integrated metadata and image management system and search gateway called NPS Focus. This system is being built with Blue Angel Enterprise software for metadata management and the LizardTech Express Server for image management. Currently ten NPS and two non-NPS databases have been integrated into the NPS Focus prototype in either full or testbed form for one-stop searching. NPS Focus has been released as an Intranet version only (http://npsfocus.nps.gov/), though a public version is projected in the near future.

Of the ten databases uploaded to date to NPS Focus, NatureBib and the NR-GIS Metadata and Data Store are most applicable to existing network bibliographic and spatial metadata. The network will continue to upload data and information to these two sites, which will coincide with the ability to search for these records through the NPS Focus portal. As NPS Focus reaches further development, other databases and repositories utilized by the network are expected to be searchable through this portal as well.

9.2.2 – Data Classification: Protected vs. Public

All data and associated information from I&M activities should be assessed to determine sensitivity. This includes, but is not limited to, reports, metadata, raw and manipulated spatial and non-spatial data, maps, etc. Network staff should carefully identify and manage any information that is considered sensitive. The Network should clearly identify and define those data needing access restrictions and those to make public.

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the National Park Service, must provide access to data and information of interest to the public. FOIA, as amended in 1996 to provide guidance for electronic information distribution, applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or the Internet. Other records are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at http://www.doi.gov/foia/ for further information.

In some cases, public access to data can be restricted according to stipulations in one Executive Order, Director's Order #66 (draft), and four resource confidentiality laws: the National Parks Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304), and the Archaeological Resources Protection Act (16 U.S.C. 470hh). Through these laws and policies, NPS is directed to protect information about the nature and location of sensitive park resources. Any information that could result in harm to natural resources can be classified as 'protected' or 'sensitive' and withheld from public release (National Parks Omnibus Management Act).

The following guidance for determining whether information should be protected is suggested in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the park resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource that is otherwise not of concern permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?

- Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

Natural Resource information that is sensitive or protected requires the:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of what data must not be released to the public
- Management and archival of those records to avoid their unintentional release

Classification of sensitive data will be the responsibility of NCCN staff, the park superintendents, and investigators working on individual projects. NCCN staff will classify sensitive data on a case by case, project by project, basis. They will work closely with investigators for each project to ensure that potentially sensitive park resources are identified, and that information about these resources is tracked throughout the project.

NCCN staff is also responsible for identifying all potentially sensitive resources to the principal investigator(s) working on each project. The investigators, whether NCCN staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases, and metadata. When submitting any products or results, investigators should specifically identify all records and other references to potentially sensitive resources. Note that partners should not release any information in a public forum before consulting with Network staff to ensure that the information is not classified as sensitive or protected.

For example, information may be withheld regarding the nature and/or specific locations of the following resources recognized as ‘sensitive’ by NPS. According to National Parks Omnibus Management Act, if NPS determines that disclosure of information would be harmful, information may be withheld concerning the nature and specific location of:

- Endangered, threatened, rare or commercially valuable National Park System resources (species and habitats)
- Mineral or paleontological objects
- Objects of cultural patrimony
- Significant caves

Note that information already in the public domain can, in general, be released to the public domain. For example, the media has reported in detail the return of condors to the Grand Canyon. If an individual requests site-specific information about where the condors have been seen, this information, in general, can be released. However, the locations of specific nest sites cannot be released.

9.2.3 – Access Restrictions on Sensitive Data

NCCN staff is responsible for managing access to sensitive data handled by the Program. All potentially sensitive park resources will be identified and investigators working on NCCN projects will be informed that:

- All data and associated information must be made available for review by NCCN staff prior to release in any format
- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service

The Network Coordinator, NPS project liaison, or Data Manager identifies all potentially sensitive park resources to the principal investigator for each project. Reciprocally, the principal investigators for each project must identify any known references to potentially sensitive park resources.

For each project, NCCN staff provides a complete list of all references to potentially sensitive park resources in each park to the park superintendent for review. Each superintendent then determines which information should be protected.

When preparing or uploading information into any NCCN database, our staff will ensure that all protected information is properly identified and marked. NCCN staff will work together to ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public media.

NCCN staff will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories developed by the I&M Program, (see section 9.2.1) are maintained on both secure and public servers, and all records that are marked 'sensitive' during uploading will only become available on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases are discussed in the guidance documents for maintaining NPSpecies and NatureBib (*in development*), as well as at the following websites:

- http://science.nature.nps.gov/im/apps/npspp/index.cfm
- http://science.nature.nps.gov/im/apps/nrbib/

Thus, access to data on sensitive park resources can be limited to Network staff or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the Network staff institute quality control and quality assurance measures to ensure that the person doing the uploading of records into the online applications is familiar with the procedures for identifying and entering protected information.

9.2.4 – Public Access to Network Inventory and Monitoring Data

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the internet. Network project data will be available to the public at one or more internet locations:

- The NCCN web site
- Public servers for the NPSpecies and NatureBib databases
- Public server for the Biodiversity Data Store
- Public server for the NR/GIS Data Store

We will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through the NCCN web site. Information on species in the national parks, including all records generated through the I&M Program, will be maintained and assessable through the NPSpecies database. Bibliographic references that refer to NPS natural resources will be accessible through the NatureBib database. Documents, maps, and data sets containing resource information from all sources, and their associated metadata, will be accessible through the Biodiversity Data Store and/or NR-GIS Data Store. Each of these databases/repositories will be available via both a secure server and a public server, and the public can access all information in these databases except those records marked as 'sensitive.'

9.2.5 – Data Availability

Both raw and manipulated data resulting from the Network's inventory and monitoring projects will be fully documented with FGDC compliant metadata and made available to the public via the Network's website. The metadata for all data sets will be made accessible to the public as soon as they are provided and verified by the Investigator(s) or project leaders.

Data sets for short-term projects will be provided to the public on the NCCN website, two years following the year the data were collected or following publication of the investigator's results (whichever comes first). Long-term (monitoring) studies will be provided to the public in four or five year intervals, or when trend analyses have been completed and reported on by the network. This will be specific to each network monitoring protocol (refer to the NCCN Monitoring Plan for further information). Before data are posted, the Investigator or project leader will be asked to verify the final data set and metadata if necessary. Once NCCN staff and the investigator verify the data set, the data will be made accessible to the public, provided no sensitive information is identified.

NCCN staff will notify investigators prior to making data sets available to the public. This will allow each investigator the opportunity to request in writing to further restrict access to the data set by the public. NCCN staff will review the Investigator's request and determine whether the request will be granted and for how long the data set will remain restricted.

9.2.6 – Data Acquisition Policy

The NCCN will develop a data set acquisition policy that will be made available to all NCCN website users who wish to acquire program data and information. This policy will include such things as:

- A mandatory questionnaire that must be completed and mailed to the project data manager before data can be acquired. This questionnaire will allow NCCN staff the ability to maintain a distribution log specifying recipient name and contact information, intended use of data, export file format, delivery date and method, and data content description noting range by date and geography of data delivered.

- A statement about use and appropriate citation of data in resulting publications.

- Request that acknowledgement be given to the National Park Service Inventory and Monitoring Program within all resulting reports and publications.

All data sets with public access available on the NCCN website will be accompanied by the acquisition policy.

9.3 – Data Feedback Mechanisms

The NCCN website will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of the NCCN I&M Program. A "comments and questions" link will be provided on the main page of the site for general questions and comments about our program and projects. A more specific "data error feedback" link will direct comments to the NCCN staff pertaining to errors found in website accessible data. Annual reporting of progress will be presented to the Board of Directors and to the Technical Committee on a yearly basis, and feedback will be expected during and following these presentations.

9.3.1 – Data Error Feedback Procedures

The following feedback procedures describe the process which NCCN will use to receive and verify data errors identified by public and private data users:

- Web users send in a notification about an alleged error through the NCCN website. NCCN staff then sends an acknowledgment to the notifier.

- NCCN staff then inputs the information into a data error log table incorporated in either each of the network monitoring databases or a specific error tracking database developed for the network.

- NCCN then determines if the data questioned by the notifier are correct or incorrect. If the data are correct, then NCCN staff informs the notifier that no corrections are to be made and the information stands. If the data are incorrect, the Network staff makes the

appropriate corrections and notifies the original data collectors (cooperator, other agency, park staff, etc…).

- Once the data are corrected, the NCCN website will be refreshed with the corrected information.

- Throughout this process, NCCN staff will continually inform the notifier via e-mail of the status.

Credits

This chapter was adapted from material developed by Sara Stevens (Northeast Coastal and Barrier Network) and Doug Wilder (Central Alaska Network).

Chapter 10 – Data Maintenance, Storage and Archiving

Effective long-term data maintenance depends on thoughtful and appropriate data documentation. An essential part of any archive is its accompanying explanatory materials (Olson and McCord 1998). This chapter will refer to, and in some cases elaborate on, metadata standards and data set documentation procedures that are more fully explained in Chapter 7 (Data Documentation).

Data, documents, and anything that results from projects and activities that use NCCN data are all crucial pieces of information. To ensure high-quality long-term management and maintenance of this information, we will implement procedures to protect information over time. These procedures will permit a broad range of users to easily obtain, share, and properly interpret both current and archived information.

10.1 – Digital Data Maintenance

In general, digital data maintained for a long time will be one of two types: short-term data sets, for which data collection and modification have been completed (i.e., inventory projects); and long-term monitoring data sets, for which data acquisition and entry will continue indefinitely.

Following the lead of the National Park Service and the National Inventory & Monitoring program, NCCN has adopted MS-Access as its database standard and ArcGIS as its spatial data management standard. NCCN will remain current and compatible with National Park Service and National Inventory and Monitoring version standards for these software programs.

Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if they are stored in out-of-date software programs or on outmoded media. Effective maintenance of digital files depends on the proper management of a continuously changing infrastructure of hardware, software, file formats, and storage media. Major changes in hardware can be expected every 1-2 years and in software every 1-5 years (Vogt-O'Connor 2000). As software and hardware evolve, data sets must be consistently migrated to new platforms or software, or they must be saved in formats that are independent of specific platforms or software (e.g., ASCII delimited files).

Any data set for which data entry or updates is still occurring will be stored under the "Active Data Projects" directory on the NCCN server. We reserve the "Project Archive" directory for data sets that will no longer change.

10.1.1 – Short-term Data Sets

Upon project finalization, project leaders are responsible for creating a set of ASCII comma-delimited text files for all short-term data sets created or managed by NCCN. Each database table will have a corresponding ASCII file. These files will be accompanied by a README.TXT file that explains the contents of each file, file relationships, and field definitions. These ASCII files are in addition to the native version of the data set (typically in database or spreadsheet format). Creating these text files will help ensure that the data are usable in a wide range of applications

or platforms. All finalized files will be stored on the NCCN server's archive section in the appropriate project folder.

In addition to creating ASCII files, NCCN will also update completed and archived data sets that may have been created in older versions of MS-Access, with the goal of having no data set more than two versions behind the current version used by NCCN. There is a risk of reducing performance in the process of conversion; for example, complex data entry forms or reports may not function properly in an upgraded version. To the extent possible, we will maintain proper functionality of data entry forms and reports. However, the priority will be to ensure basic table and relationship integrity. All previous versions of the data set will be saved.

10.1.2 – Long-term Monitoring Data Sets

Long-term monitoring data sets require regular updates and conversion to current database formats. All active or long-term databases will conform to the current NPS and I&M software version standards.

Monitoring projects will also have variable long-term data archiving requirements. Raw data sets that are later manipulated or synthesized may need to be stored in perpetuity. Modifications to protocols will typically require complete data sets to be archived before modifications are implemented. Depending on the monitoring project, it may be necessary to preserve interim data sets (data "milestones") over the long term. We will save archived data sets or subsets destined for long-term archiving, whenever possible, in their native formats in addition to ASCII text files. Data archiving requirements for ongoing projects will be spelled out in the data management SOPs for each monitoring project.

10.1.3 – Quality Control of Converted Data

All ASCII files created from databases will undergo Quality Control (QC) to ensure that the number of records and fields correspond to the source data set, and that conversion has not created errors or data loss. A second reviewer (preferably a program scientist) will evaluate the ASCII files and documentation to verify that tables, fields, and relationships are fully explained and presented in a way that is useful to secondary users.

Databases that are converted from one version of MS-Access to an upgraded version will require additional QC. This is especially important if the databases are being actively used for data entry or analysis. Forms, queries, reports, and data entry all will be thoroughly tested.

10.1.4 – Version Control

Previous versions of databases will be saved in their native format and archived in addition to the current version. Documentation of version updates and associated details will be part of the archive metadata document, and revision information and history will also be included in tables within the database files themselves. File names of the archived revisions should clearly indicate the revision number or date.

10.1.5 – Spatial Data

Spatial data sets that are essential to NCCN will be maintained in a format that remains fully-accessible by the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats, and while shapefiles have retained all functionality in ArcGIS, coverages may require conversion to ArcGIS format if they are no longer supported. At this time there is no practical way to save GIS data in a software- or platform-independent format.

Both uncorrected and corrected GPS data (e.g., .SSF and .COR files) will be archived in their native format in addition to the corresponding GIS files that are created.

10.2 – Storage and Archiving Procedures – Digital Data

Digital data will be stored in a repository that ensures security and ready access to the data in perpetuity.

10.2.1 – Directory Structure for Digital Archives

Figure 10.1 illustrates an example of a directory structure for both active and archived data files, including GIS data. This directory structure was devised, in part, to accommodate two different backup schedules: one for rapidly-changing files; the other for relatively static files.

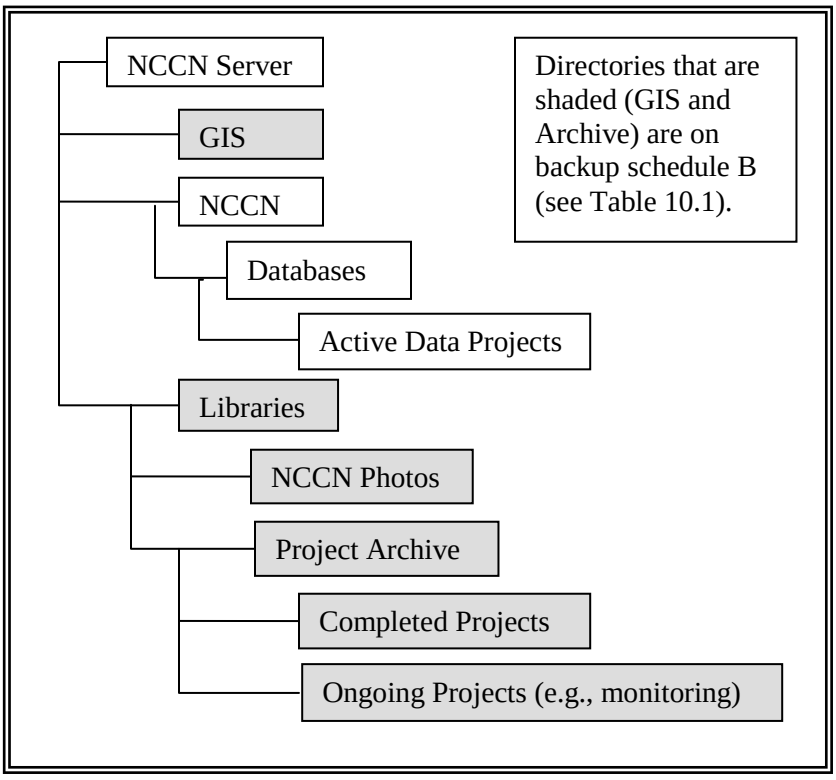

Figure 10.1. Schematic of proposed data components of the NCCN servers

10.2.2 – Directory Structure for Individual Projects

The organizing and naming of folders and files should be intuitive enough that users unfamiliar with a specific project can still easily navigate it. Since each project has its own variations and idiosyncrasies, a standardized structure isn't realistic. However, all project archives will include several to most of the following elements:

- administrative documents such as agreements, correspondence, and research permits
- programmatic documents including protocols, procedures, and supporting documents
- interim data sets or milestones
- data sets submitted by contractors
- data sets reformatted or manipulated by NCCN (e.g., data converted to NRDT format, data sets migrated to current software formats)
- data sets in ASCII format
- conceptual or statistical models used for data interpretation
- final reports
- README files, including an explanation of directory contents, project metadata, and version documentation

Once final data and reports have been submitted, draft products do not need to be maintained.

10.2.3 – Backup Procedures for Digital Data

The risk of data loss can come from a variety of sources, including catastrophic events (e.g., fire, flood), user error, hardware failure, software failure or corruption, and intentional acts of vandalism. Performing regular backups of data and arranging for off-site storage of backup data sets are the most important safeguards against data loss.

NCCN is integrated into the NPS Pacific West Region (PWR) wide area network and benefits from the interconnectivity of the PWR system. Data residing on one of the NCCN servers are backed up to another NCCN server using an automated backup system (see Figure 3.5). This integrated system is in development and final backup schedules will be included as appendices to this DMP and will be developed in cooperation with IT system administrators. Currently each NPS unit functions independently with backups done locally and stored off site.

Table 10.1. Proposed backup schedule for NCCN servers

Schedule A – frequently-changing files; estimate total of 50-75 GB by 2006. Used for all files located on NCCN server	
Tape Set 1	week 1: full backup and daily incrementals week 2: daily incrementals week 3: daily incrementals week 4: daily incrementals – rotate off-site
Tape Set 2	week 5: full backup and daily incrementals week 6: daily incrementals week 7: daily incrementals week 8: daily incrementals – rotate off-site
Tape Set 3	week 9: full backup and daily incrementals week 10: daily incrementals week 11: daily incrementals week 12: daily incrementals – rotate off-site
Schedule B – large files, relatively static; estimate total of 350GB by 2006. Used for all files located on NCCN server\GIS and NCCN server\Archive	
Tape Set 1	week 1: full backup weeks 2-26: weekly differential backup. Rotate off-site at week 26.
Tape Set 2	week 27: full backup weeks 28-52: weekly differential backup. Rotate off-site at week 52.
Tape Set 3	week 1 (year 2): full backup weeks 2-26 (year 2): weekly differential backup. Rotate off-site at week 27.

Backups of data stored on the personal computers of staff are the responsibility of each staff member. NCCN recommends that staff regularly copy important files onto a personal directory on the local-area network, where daily backups will be performed. Staff may also place data in the archive section of the NCCN server with the approval of the data manager.

The quantity of NCCN data is increasing rapidly. Thus, we will need to reevaluate the backup schedule often, and at least annually. Backup routines represent a significant investment in hardware, media, and staff time; however, they are just a small percentage of the overall investment that we make in program data.

10.2.4 – Testing Backup Files

All backups should be run with "verify" toggled on to compare source files against the backup files and detect any errors and discrepancies. However, to protect against hardware failures, a RAID-5 array with multiple hard drives for data storage and ArcServe for backing up to these hard drives will provide redundancy that prevents data loss in the event of hard drive failure. Periodic test restoration of a few files will assure that the system works and keep administrators refreshed in technique.

10.2.5 – Data and Network Security

NCCN local and wide area networks currently conform to Department of Interior security guidelines. Only NCCN park and network staff and system administrators have permission to

access files on the NCCN servers, and restrictions have been established on archived data files. Directories containing completed project data or interim versions of ongoing projects are designated as read-only for all staff with the exception of the data managers. Therefore, any changes must be routed through the data manager, who is responsible for ensuring that documentation and README files associated with the data set are also updated.

10.3 – Storage and Archiving Procedures – Documents and Objects

The guidelines in this section apply to documents such as final reports prepared by staff or contractors, program administrative documents, contracts and agreements, memoranda of agreement, and other documents related to NCCN administration, activities and projects. These guidelines also apply to physical items such as natural history specimens, photographs, or audio tapes. In most instances, these documents and objects are essential companions to the digital data archives described earlier.

Direction for managing these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records. NPS-19 further states:

> "Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record."

Section N of Appendix B, which provides guidelines on natural resource-related records (including, specifically, the results of Inventory and Monitoring Programs), indicates that all natural resource records are considered "permanent," that is, are to be transferred to the National Archives when 30 years old. It also indicates that non-archival copies of natural resource-related materials are potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed.

10.3.1 – Documents

All paper documents managed or produced by the NCCN will be housed in one of the secure repositories located in parks within the Network. Digital files are archived on the NCCN servers. These may include project files, administrative documents and non-record copies of documents that are archived at an off-site facility. Examples include: meeting minutes, correspondence, memoranda of understanding, contracts and agreements, research permits, and interim and selected final reports produced by the program or under its auspices. NCCN will use acid-free paper and folders for all permanent records in the central files. In addition to maintaining these paper records, NCCN will maintain electronic versions on the NCCN servers whenever possible. The central files are maintained by an NCCN administrative clerk, under the guidance of the data managers and network coordinator.

An archival repository provides temperature and humidity-controlled facilities, a professional archival staff, and meets all museum standards set by NPS. These repositories will be used for original documents and associated materials produced by the network (e.g., photographs, field notes, specimens, permits) that are a high priority to maintain under archival conditions. Examples include: original inventory reports and accompanying slides and maps; original vegetation mapping reports; NCCN Implementation Plan (Phase 1, 2 and 3 reports). Copies of these reports will be maintained in the NCCN I&M central files, and all will have an electronic equivalent (e.g., PDF) for distribution or reproduction.

For all materials submitted to an archive, the submitter will provide essential cataloging information such as the scope of content, project purpose, and non-record copies of documents that are presented using archival-quality materials (acid-free paper and folders, polypropylene or polyethylene slide pages).

Many NCCN reports and documents encompass data from multiple parks, which has made it difficult to accession archival copies into a specific network park museum. In these instances NPS PWR Seattle Library will catalog records that reference parks included in a report or document and will prepare finding aids to help potential users locate the materials.

NCCN will provide high-quality copies of park-related documents resulting from NCCN projects, along with electronic versions, to park resource management staff. Parks may choose to accession these materials into their museums, incorporate them into their central files, or house them in their resource management library as they deem appropriate. NCCN will not manage documents at the Park level.

10.3.2 – Specimens

The Network will provide specimens collected under the auspices of the NCCN to the Network Park in which they were collected, or to a repository approved by a Park (where the specimens are considered on loan). NCCN will provide Park curators with associated data required for cataloging each specimen. These data will be in comma-delimited format (.CSV) format for automated uploading into ANCS+. Data will be provided to non-NPS curators in Microsoft Excel format.

10.3.3 – Photographs

Archivists have been reluctant to fully embrace digital photography, and some have expressed concern that, with the acceleration of technological change, documentary heritage is in danger of being lost in the information age (Cox 2000).

NCCN has chosen to take a conservative approach and requires staff and contractors to provide photos as 35mm slides (preferably Kodachrome or Ektachrome), which have a proven long-term stability (Wilhelm and Brower 1993). If contractors cannot provide slides, NCCN requests 4x6 color prints. Original images are a high priority for placing in archival storage conditions.

Slides should be labeled using indelible pigment ink or using laser-printed, archival-quality slide labels. Slide labels should include: a unique ID, project name, photographer, photo date, a brief identification of contents (e.g., species name, plot ID), and geographic location (UTMs or description). All slides will be stored in polypropylene slide sleeves at the park office until transferred to the appropriate long-term curation facility. In addition, all slides will be scanned and saved as TIFF files, and these electronic copies will be used as the primary means of distributing or reproducing the images.

If photographs are provided, they will be stored in individual polypropylene sleeves and within archival boxes. Each photo will be labeled on the back, using archival-quality labels that are either laser-printed or hand-labeled, with the same information elements required for slides. If a contractor is submitting photographs, corresponding digital images in TIFF format should also be submitted

Every image, regardless of format, will have an entry in the NCCN Photo Database (*guidance documents in development*), where attributes such as electronic file name, keywords, project, photo description, photographer, date, and location will be catalogued. All photo files and the associated photo database are housed on the archive portion of the NCCN server (see Figure 10.1).

Role of Curators in Storage and Archiving Procedures

Curators for parks within NCCN are an ongoing source of expertise, advice, and guidance on archiving and curatorial issues, and they have a role in almost every project undertaken by the Network. Project leaders should involve Park curators when projects are in the planning stage, to ensure that all aspects of specimen preservation or document archiving will be considered and that any associated expenses are included in project budgets.

Credits

This chapter was adapted from material developed by Margaret Beer (Northern Colorado Plateau Network).

Glossary

Base cartographic data – standard map layers such as boundaries, roads, trails, hydrography; typically not project-specific

EBLA – Ebey's Landing National Historical Reserve, www.nps.gov/ebla

FGDC – Federal Geographic Data Committee determines standards for geospatial data

FOVA – Fort Vancouver National Historic Site, www.nps.gov/fova

GIS – Geographic Information Systems, software used to analyze spatial data and create maps

Guidelines – general descriptions of data management practices and procedures

I&M – the NPS Inventory and Monitoring Program: http://science.nature.nps.gov/im/index.cfm

IAR – Investigator's Annual Report, turned in each year by principal investigator; describes work accomplished during preceding year on a specific project

IT – information technology, composed of hardware, software, digital media, computer and networking services, plus the personnel and knowledge base needed to maintain them

Inventory – "An extensive point-in-time effort to determine location or condition of a resource, including the presence, class, distribution, and status of plants, animals, and abiotic components such as water, soils, landforms, and climate. Inventories contribute to a statement of park resources, which is best described in relation to a standard condition such as the natural or unimpaired state. Inventories may involve both the compilation of existing information and the acquisition of new information. They may be relative to either a particular point in space (synoptic) or time (temporal)." (Source: http://science.nature.nps.gov/im/monitor/ProgramGoals.cfm#Definitions)

Layer – a GIS digital equivalent of a mylar map overlay made up of points, lines or polygons

LEWI – Lewis and Clark National Historical Park, www.nps.gov/lewi

Local Area Network (LAN) – an interconnected system of computers connected to one or more network servers by cabling and administered as a collective unit

Long-term monitoring project – a monitoring study designed to assess the status of a natural resource vital sign over time

Metadata – information about a data set that tells the who, why, when, where and how the data was collected

MORA – Mount Rainier National Park, www.nps.gov/mora

NBII – National Biological Information Infrastructure, a system of standards for describing biological data

Network – in the context of the NPS Inventory and Monitoring program, monitoring networks are groups of parks linked by geography and shared natural resource characteristics. The network arrangement is intended to provide a minimum infrastructure for initiating natural resource monitoring in the approximately 270 parks with significant natural resources.

NOCA – North Cascades National Park Service Complex, www.nps.gov/noca

OLYM – Olympic National Park, www.nps.gov/olym

Pick list – a drop-down list of potential values to choose from to enter data into a table or form

Primary key – a unique identifier field in a database table that can link records in different tables to each other

Protocol – a set of sampling procedures and/or experimental processes followed in inventory studies or pilot monitoring projects. In the context of the Inventory and Monitoring Program, this often indicates a monitoring protocol, which is the set of formal documents and sampling processes describing how a vital sign will be monitored. Monitoring protocols are composed of a narrative section, standard operating procedures and supplementary information (databases, reports, tools, hardcopy materials)

Prototype – in the context of the NPS Inventory and Monitoring program, prototype programs play a special role as centers of excellence and as sources of technical guidance and expertise for other monitoring programs across NPS.

Query – an object created in a relational database that extracts selected data from table(s)

Referential integrity – where cascade updates to related fields and cascade deletes to related records are maintained

Relational database – a database with data stored in multiple tables where each table contains data on one subject and tables are linked by key fields to link records in different tables

Relationship – in a relational database it is the way that records are linked between tables; can be one to one, one to many or many to many

RPRS – Research Permit Reporting System for tracking research conducted in NPS units

SAJH – San Juan Island National Historical Park, www.nps.gov/sajh

Standard Operating Procedures (SOPs) - detailed step-by-step instructions for carrying out sampling procedures in monitoring Protocols; dynamic in nature and frequently updated; can reference overall data management plan, standards and DMOPs

Standard - a data management practice or policy implemented Service-wide or network-wide (*e.g.,* recommended naming conventions, FGDC metadata format); referenced by overall data management plan, monitoring SOPs and DMOPs

SQL – Structured Query Language used to write statements used by software programs to manipulate data; also the name of a software program that runs on a server that makes data available to clients

Theme – outdated name for GIS layer made up of points, lines or polygons for mapping features

WASO – Washington Support Office

Wide Area Network (WAN) – an interconnected system of local area networks servers connected to each other by phone lines or dedicated server lines; administration decentralized

References

Chapal, Scott E., and Don Edwards. 1994. Automated smoothing techniques for visualization and quality control of long-term environmental data. Pp. 141-158 *in* Michener, William K., James W. Brunt, and Susan G. Stafford (eds), Environmental Information Management and Analysis: Ecosystem to Global Scales. London: Taylor & Francis Ltd.

Cox, R. J. 2000. Searching for authority: archivists and electronic records in the New World at the fin-de-siècle. First Monday 5:1. http://firstmonday.org/issues/issue5_1/cox/index.html

Edwards, Don. 2000. Data quality assurance. Pp. 70-91 *in* Michener, William K., and James W. Brunt (eds), Ecological Data: Design, Management and Processing, Methods in Ecology Series, Malden, MA: Blackwell Science Ltd.

Michener, William K., and James W. Brunt (eds). 2000. Ecological Data: Design, Management and Processing, Methods in Ecology Series, Malden, MA: Blackwell Science Ltd.

Olson, R. J. and R. A. McCord. 1998. Data Archival. In: Michener, W.K., J.H. Porter, and S.G. Stafford. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque, NM. pp 53-57.

Palmer, Craig J., and Eric. B. Landis. 2002 Draft. Lake Mead National Recreational Area Resource Management Division: Quality System Management Plan for Environmental Data Collection Projects.

Tessler, Steve, and Joe Gregson. 1997. Draft Data Management Protocol. NPS I&M Program. http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm

USDI National Park Service. 2003. NPS records disposition schedule. NPS-19, Appendix B – Revised, 5-03. 61 p.

U.S. Environmental Protection Agency. 2003. Guidance for geospatial data quality assurance project plans. EPA QA/G-5G. EPA/240/R-03/003. Washington, DC: Office of Environmental Information.

Vogt-O'Connor, D. 1997. Caring for photographs: general guidelines. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC. 4 p.

Vogt-O'Connor, D. 2000. Planning digital projects for preservation and access. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC. 4 p.

Wilhelm, H. and C. Brower. 1993. The permanence and care of color photographs: traditional and digital color prints, color negatives, slides, and motion pictures. Preservation Publishing Company, Grinnel, IA. 744 p.

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS D-48, February 2009

National Park Service
U.S. Department of the Interior

Natural Resource Program Center
1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525

www.nature.nps.gov